ENHANCING H.26X CODING FOR VISUAL COMMUNICATIONS - WITH APPLICATIONS IN TELEMEDICINE AND TELEVISION

A Thesis
Presented to
The Academic Faculty

by

Sourabh M. Khire

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
May 2013

ENHANCING H.26X CODING FOR VISUAL COMMUNICATIONS - WITH APPLICATIONS IN TELEMEDICINE AND TELEVISION

Approved by:

Professor Nikil Jayant, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Ghassan Al-Regib
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Raghupathy Sivakumar
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Lee Cooper
Department of Biomedical Informatics
Emory University School of Medicine

Professor David Anderson
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Arturo Rodriguez
Video Technology Group
Cisco Systems

Date Approved: 25 February 2013

To my family, and the beautiful sport of tennis.

ACKNOWLEDGEMENTS

During my time at Georgia Tech I have had the pleasure to learn, collaborate and meet many a wonderful people. I would like to thank each and every one of them for contributing in the making of this thesis. In particular, I would like to thank Dr. Nikil Jayant for being such a brilliant advisor. I would like to thank Dr. Jayant for allowing me the freedom to pursue research in the areas of my liking, and guiding me through this thesis with his unique brand of patient and supportive mentoring. I would also like to express my gratitude to Dr. Jayant for teaching by example, the value of professionalism and good communication. These are lessons I will carry with me through the rest of my professional and personal life. I am truly grateful to Dr. Arturo Rodriguez for sharing his unique industry perspective with me, which helped shape a major part of my thesis. I would also like to thank Dr. Raghupathy Sivakumar, Dr. David Anderson, Dr. Ghassan Al-Regib and Dr. Lee Cooper for agreeing to serve on my thesis committee, and for taking the time to read my thesis and attend my defense.

It has been an absolute privilege to collaborate on various projects with Lee Cooper and Alexis Carter from Emory University, Saunya Williams, Eun Seok Ryu, and Scott Robertson from Georgia Tech, Max Stachura and Elena Wood from the Georgia Regents University, Arturo Rodriguez and Anil Katti from Cisco Systems, and Col. Tamer Goksel. I would like to thank all these collaborators for their valuable contributions to the research presented in this thesis. Obtaining a degree from Georgia Tech involves a lot of courses and research, but also involves filling a lot of forms, compiling a lot of resources, and scheduling and meeting a lot of deadlines. Fortunately, none of these administrative burdens ever hindered my research progress

thanks to the wonderful staff at GCATT comprising of Barbara, Tina, Kevin, Rex and JoAnna. I would especially like to thank Barbara, for all that she has done for me, and for other students at the MMC lab over the years. I would also like to thank Jeannie, Rama, Saunya and Shira for helping me through my research, and for making the time spent in the lab a pleasant experience.

I have been extremely fortunate to have met a great bunch of friends at work, and away from work over the past few years. Life at Georgia Tech to me has been akin to a long tennis match full of highlight reels, a few shanked forehands and some missed calls and unlucky net-cords thrown in for good measure. Given the space limitations of this document, it is impossible to acknowledge everybody by name, but I would like to extend big thanks to each and every one of my friends for helping me survive through the ebb and flows of this long five setter. Finally, I would like to thank my parents Jayashree and Mohan, and my sister Sujata for their loving support over the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Video as a visual-communication modality finds utility in multimedia applications ranging from entertainment and education to telemedicine and telehealth. However, video delivery for enabling such a multitude of applications imposes substantial demands on the underlying communication infrastructure. These demands are further amplified with the advent of mobile multimedia technology, which while simplifying the production of high-quality videos, also multiplies the challenges in providing a high-quality video experience to the end user. In a wireless and mobile communication paradigm, distribution and sharing of video content often occurs over unfriendly network environments constrained by lack of sufficient bandwidth, and prone to jitter, delay and packet losses. The research presented in this thesis reasons that it is not possible to design a single "silver bullet" video-coding and video-delivery solution to address all the challenges posed by the variety of multimedia applications that need to be supported over a diverse set of channel conditions. Instead, this research proposes an assortment of application-specific optimizations designed to enable high-quality video communication over bandwidth constrained and unreliable channels, termed herein as Application Specific Video Coding and Delivery (ASVCD) solutions. These ASVCD solutions employ content and context adaptive approaches such as Region of Interest (ROI) video coding, Multiple Representation Coding (MRC), and Multiple Representation Coding of the Region of Interest (ROI + MRC) for enabling video-based applications such as live surgical telementoring, mobile and wireless video streaming, and asynchronous distance learning.

Stated briefly, this thesis investigates the effectiveness of ROI based video-coding in facilitating diagnostically lossless delivery of surgical videos, with the objective of

enabling surgical telementoring over very low bandwidth channels. Furthermore, to facilitate error resilient video delivery over channels prone to burst losses and signal loss intervals, a novel scheme known as Multiple Representation Coding is proposed and evaluated in this thesis. Finally, the thesis proposes a scheme for unequal protection of the ROI in the video by using the MRC scheme to enable applications such as distance learning and video streaming over low-bandwidth and unreliable channels. The ultimate aim of every application-specific optimization presented in this thesis is to provide the end-user with a high-quality multimedia experience in spite of the limitations of, and the impairments on the underlying network.

CHAPTER I

INTRODUCTION

The relative ease in producing, sharing and consuming video content afforded by the recent advances in telecommunication and silicon technology, has multiplied the number of multimedia applications and services available to the average consumer. Today, viewers can receive television broadcasts using Internet Protocol Television (IPTV), stream on-demand videos using Netflix and Hulu, and engage in virtual meetings and video conferences using Cisco Telepresence and Skype. Additionally, due to the business growth reported by some e-retailers [82], availability of product-specific videos on e-commerce websites such as Amazon and e-Bay is also becoming increasingly common. This explosion in the popularity of streaming video is illustrated in studies which have reported that over 85% percent of the Internet audience in the United States views online videos every month [21], and that billions of videos are consumed globally in a month [20]. Besides entertainment and shopping, video communication plays a crucial role in facilitating important applications such as telemedicine and remote healthcare delivery. Delivering medical videos with a diagnostically lossless (DL) quality can lower morbidity by enabling a variety of telemedicine applications such as teleconsultation, telementoring, telesurgery, and remote diagnosis. For example, telementoring has already been proven effective in multiple scenarios ranging from surgical education and civilian-oriented peace-time health care, to emergency care in disaster relief and in military combat zones [29, 78, 83].

An important factor contributing to the exponential growth in the availability of the so-called "rich media applications" has been the improvement in the video-delivery systems, and the video-coding technology facilitated by the H.26x family of

video-coding standards. The H.26x series [102] comprises of the H.261 [37] and the H.263 [38] standards developed by the ITU-T Video Coding Experts Group (VCEG), and also the H.262/MPEG-2 Video [68], the H.264/MPEG-4 Advanced Video Coding (AVC) [39], and the H.265/High Efficiency Video Coding (HEVC) [13] standards jointly produced by the ITU-T/VCEG and the ISO/IEC Moving Picture Experts Group (MPEG). H.264/AVC is currently the most deployed video compression standard, and H.265/HEVC is the new emerging standard, promising a bit-rate reduction of around 50% for equal visual quality compared to the H.264/AVC video-coding standard. In spite of the widespread deployment of the H.26x series of video encoders, it is important to note that the different users participating in video communication over the Internet are equipped with devices differing in power and computational capabilities, and communicate over networks with differing bounds on the bandwidth and channel reliability. For example, a video uploaded on a globally visible video-sharing website such as Youtube may be streamed by a user in an urban setting equipped with a fairly powerful desktop computer, and having access to a broadband connection with a download speed of several megabits per second (Mbps). The same video could also be requested by user located in a remote rural area over a cellular data connection with a download speed of only a few hundred kilobits per second (kbps). In addition to such client diversity, the multimedia applications themselves impose substantial and diverse set of demands on the underlying network. It is not possible to design a single video coding and video delivery solution that can effectively address the enormous diversity and heterogeneity created by variables such as channel conditions and computing limitations, all while satisfying the unique requirements of each and every multimedia application that needs to be supported. Clearly, enabling high-quality video communication over bandwidth constrained and/or unreliable channels requires certain application-specific optimizations.

In this thesis we present several video-coding and video-delivery solutions which

adapt to the nature of the underlying network, and also to the modality of the video content provisioned by each application. Collectively, we refer to this assortment of content and network adaptive algorithms as Application Specific Video Coding and Delivery solutions. The rest of this chapter will be devoted to establishing the need for developing ASVCD solutions, and to introducing the reader to example applications which can be rendered effective by employing the ASVCD toolkit.

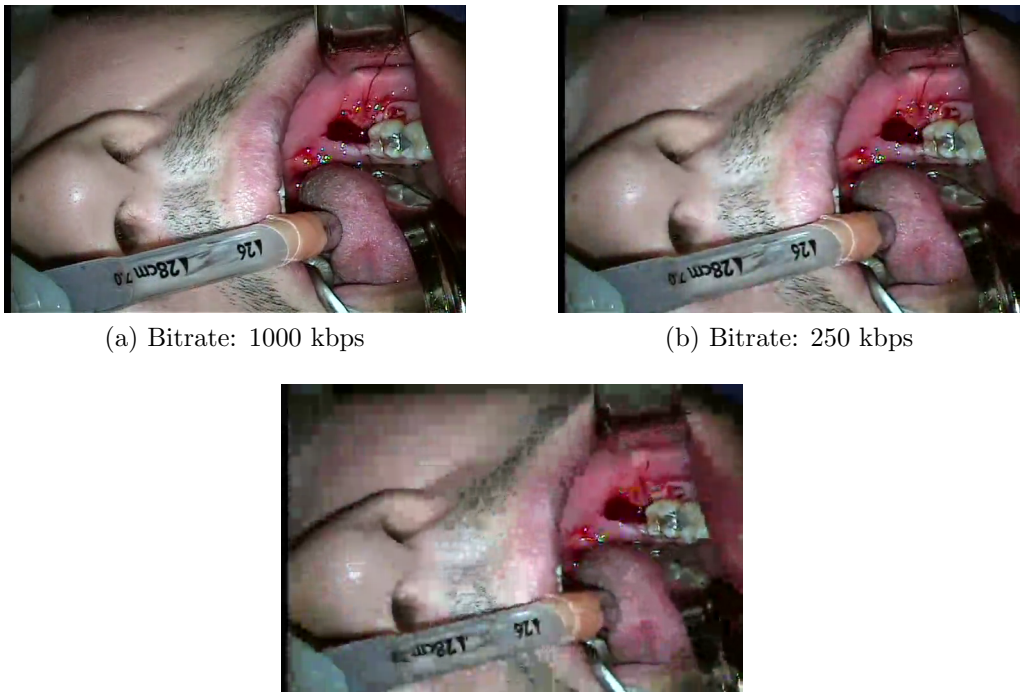
Diagnostically lossless delivery of surgical procedures over low-bandwidth channels

Surgical telementoring can be very effective in overcoming distance, cost, and mobility issues, but the necessary bandwidth availability is not universal. Scenarios requiring ultra-low bandwidth video communication are very likely to occur in rural or military settings. Existing compression standards such as H.264/MPEG-4 Advanced Video Coding (AVC) [39] can compress videos at very low bitrates (less than 200kbps), but the quality of the compressed video is almost never "diagnostically lossless." Thus, telementoring in low-bandwidth scenarios is an example of applications that could benefit from the improved end-to-end video delivery facilitated by the proposed ASVCD solutions. A compressed medical image or video is said to be diagnostically lossless, if the medical expert examining the medical imagery does not feel uncomfortable rendering a diagnosis off it. Based on expert evaluations it was determined that the DL compression ratio for digital pathology images compressed using the Joint Photographic Experts Group (JPEG) standard [49] is approximately 10:1 [109], and that videos of laproscopic surgery also compressed using JPEG are diagnostically lossless at bitrates higher than 1.5 Mbps [42]. Compressing the medical imagery at DL or sub-DL compression ratios is crucial to ensure that the information conveyed by important features such as the myocardium in a cardiac MRI [19], the nuclear morphometry, intensity, texture and gradient information in stains of

Gliomas [110], or medical features such as skin tone mottling and respiratory rate in case of telepediatric videos [81] is not compromised. However, aggressive compression of image or video data can result in artifacts that can impair the visibility of diagnostically important features. For example, as seen in Figure 1, the visibility of the suture degrades as the encoded bitrate of the surgical video drops from 1000 kbps (suture clearly visible) down to 250 kbps (suture visibility affected), and to 100 kbps (suture is not at all visible). Unfortunately, sufficient bandwidth to encode and transmit the entire video at a DL bitrate is often unavailable. In this thesis, we recommend employing a content-adaptive video-coding solution, known as Region of Interest compression to overcome this challenge. In ROI-based video encoders, the diagnostically or medically important regions of the frame are compressed at a higher bitrate, and hence at a higher quality as compared to the rest of the frame. ROI video coding, thus enables compression and transmission of the medical features at a diagnostically lossless quality, without exceeding the bit budget dictated by the channel bandwidth. Hence ROI-based video coding is included as one possible video-coding method in the ASVCD suite of solutions.

Error-resilient video delivery in the presence of burst losses and signal loss intervals

Besides bandwidth constraints, wireless and mobile video access is also hampered by the presence of burst losses and signal-loss intervals over the network. For example, users may experience a complete loss of signal when riding an elevator, driving through a tunnel, or traveling aboard an underground public transit. Signal loss on the physical layer results in the loss of multiple packets on the network layer. Loss of several packets ultimately manifest as a loss of multiple frames or seconds of the video sequence on the application (video) layer. Existing error-control and error-recovery



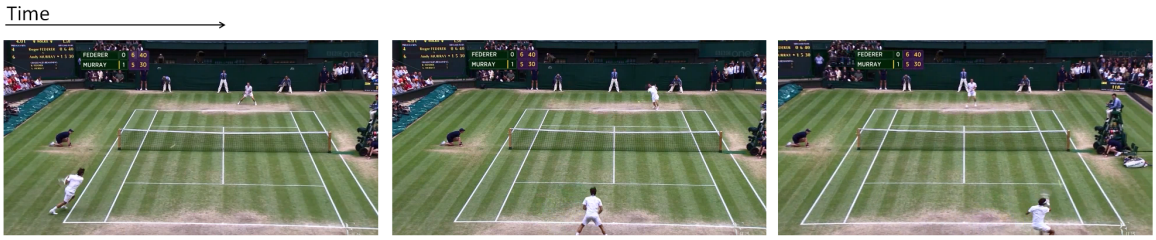
(a) Bitrate: 1000 kbps

(b) Bitrate: 250 kbps

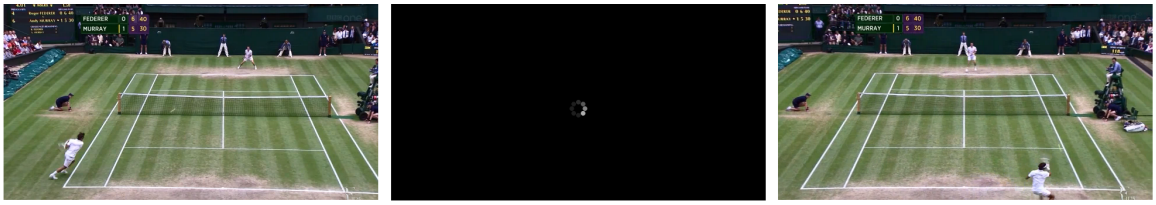
(c) Bitrate: 100 kbps

Figure 1: Effect of bitrate on the suture visibility in a surgical video.

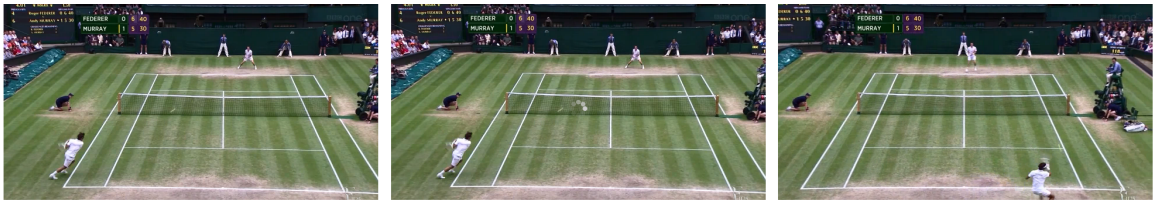
approaches such as Forward Error Correction (FEC), or Flexible Macroblock Ordering (FMO) defined in the H.264/AVC standard cannot handle multi-frame losses, and hence are rather limited in the presence of burst errors or signal-loss intervals. As a result, the decoder has to resort to some form of naive error-concealment if multiple consecutive frames of the sequence are lost. This is illustrated in Figure 2, where a loss of multiple consecutive frames results in a "black-screen" or a "freeze frame" at the receiver. This severely hampers the quality of video experience for the end-user. Objectively, the loss of quality due to a freeze frame or a black screen appears as deep "valleys" in the Peak Signal to Noise Ratio (PSNR), as seen in Figure 3. In this research, we propose and demonstrate the effectiveness of a novel approach for error-resilient video delivery, which we term as Multiple Representation Coding. By transmitting multiple interleaved representations of the source sequence, the MRC scheme facilitates a graceful recovery from burst and signal losses over the network.



(a) No loss



(b) Loss: black screen



(c) Loss: freeze frame

Figure 2: Effect of the loss of consecutive frames on the received sequence.

Further, we improve the error-robustness of the proposed scheme by modifying the MRC scheme to adapt to the expected length of the burst loss over the network. MRC thus constitutes one of the video-delivery methods included in the ASVCD toolkit.

Error-resilient delivery of the region of interest over bandwidth-constrained and unreliable networks

Bandwidth constraints and lossy channels can independently limit the ability to enable high-quality video communication. However, the two impairments are not necessarily mutually exclusive. On the contrary, wireless and cellular connections can often be constrained by lack of sufficient bandwidth, and simultaneously be impaired by burst losses and signal loss intervals. Useful applications such as distance learning, where students remotely participate in courses by streaming live or pre-recorded video

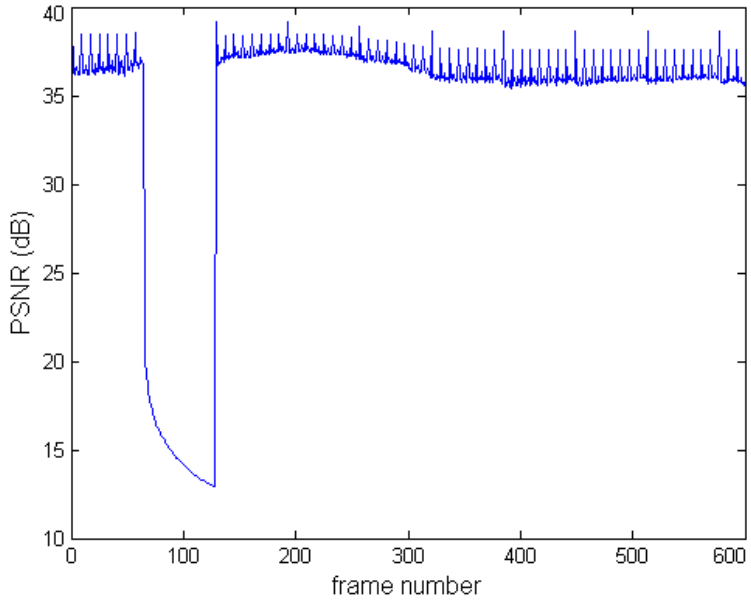


Figure 3: Effect of the loss of consecutive frames of the video on the PSNR of a coded video sequence.

lectures can be rendered undeliverable due to such impairments. We observed that videos of lectures or presentations usually contain a well-defined ROI such as an equation on the whiteboard, or the face and gestures of the lecturer. In such distance learning applications, it can be useful to employ ROI-based video coding to deliver a high-quality ROI in spite of the limited bandwidth availability. Further, error-free delivery of the ROI can be ensured by using the MRC scheme. Thus, in this thesis we propose disproportionately allocating the bits in a frame to establish a high-quality ROI, and then unequally protecting the ROI against burst and signal losses by using the MRC scheme. This coupling of video-coding with a ROI support, and error-resilient video delivery using MRC is also included in the suite of ASVCD solutions.

1.1 Key Contributions

Based on the discussion above, we can state that existing line of H.26x video-coding standards although powerful, still need to be adapted to provide higher coding efficiency and reliable video delivery for enabling several important applications. Further, we believe that these optimizations would need to be both content dependent

and network dependent. Supported by multiple research grants, and strengthened by a knowledge base adapted from signal processing and communications theory, this thesis makes three key contributions to facilitate high-quality video communication tailored to support specific multimedia applications over non-ideal networks.

- We investigate the effectiveness of ROI video coding in enabling surgical telementoring over very low-bandwidth channels. We implement a low-complexity, standard-compliant, flexible and interactive ROI support in surgical videos, and gather expert feedback to evaluate the diagnostic quality of the surgical field when encoded as the ROI. Further, we also develop a working prototype demonstrating a surgical telementoring application enabled using the proposed ROI video coding technology. Finally, we compare the performance of ROI-based video coding with other techniques such frame-rate and frame-resolution downsampling, which may also be useful in diagnostically lossless delivery of medical videos.

- We present a novel approach for enabling error-resilient mobile and wireless video streaming, which we term as Multiple Representation Coding. We demonstrate that the MRC scheme can facilitate a graceful recovery from impairments on the coded video sequence due to burst losses or signal loss intervals. Further, we enhance the performance of the MRC scheme by implementing modifications which enable the MRC scheme to adapt to the length of the burst loss over the underlying network. We also compare the performance of the MRC scheme with FEC, and study the effectiveness of the MRC scheme in enabling error-resilient delivery of sequences encoded using the emerging H.265/High Efficiency Video Coding (HEVC) standard.

- Finally, we present the ROI + MRC method for facilitating robust delivery of a high-quality ROI by unequal protecting the ROI using the MRC scheme, at the

expense of a degraded and lossy background. The ROI+MRC technique will be employed for enabling applications such as asynchronous distance learning.

1.2 Organization of the thesis

The three key contributions of the thesis described in the previous section constitute the ASVCD toolkit, and will be described in the upcoming three chapters. It is important to note that each of these techniques have a "sweet spot" in terms of possible application scenarios. For example, ROI-based video coding is most effective in enabling applications which require real-time video communication over low-bandwidth channels. The MRC technique owing to its design, is most suitable for enabling non real-time applications such as video streaming which require error-resilient video delivery. Since the ROI+MRC technique inherits the advantages and limitations of both methods, the ROI+MRC technique is most effective in enabling video applications which are hampered by the presence of bandwidth constraints and channel impairments. Consequently, each upcoming chapter presents the problem definition, an overview of existing research in each area, and also the specific application scenarios where the ASVCD toolkit finds the most utility. Further, a discussion of the experimental results, and the conclusion derived from engaging in this research is also documented in each chapter. Thus, Chapter 2 presents on overview of existing telementoring systems, and ROI video coding methods. Implementation and evaluation of the proposed ROI-based video coding solution is also presented in Chapter 2. A literature review of existing error-resiliency schemes for video delivery is presented in Chapter 3. The MRC and the adaptive-MRC schemes are introduced, and compared to an existing scheme in Chapter 3. Chapter 4 presents techniques to enable a robust delivery of the ROI, with the aim of enabling distance learning applications. Finally, Chapter 5 summarizes the major take-away from this research, and presents avenues for future research.

CHAPTER II

REGION OF INTEREST VIDEO CODING FOR ENABLING SURGICAL TELEMENTORING IN LOW-BANDWIDTH SCENARIOS

2.1 Introduction

Surgical needs of remote rural communities and those of the armed forces present a unique set of challenges. In rural surgical facilities or military field hospitals, a single surgeon may oversee the care of numerous patients. However given the extent of specialization in surgery, just one surgeon cannot be expected to handle the wide range of surgical possibilities that may arise in such settings. Surgical telementoring presents a possible solution to this problem. Surgical telementoring represents an advanced telemedicine application which can facilitate real-time guidance and instruction of less experienced, remotely-located general surgeons by more-experienced surgeons and subspecialists using audio, video, and telecommunication resources. It has been shown that surgical telementoring can increase patient access to experienced surgeons and expert consultation in rural communities, and facilitate emergency care during disaster relief, humanitarian missions, and in military combat zones [29,78,83]. Further, surgical telementoring can prove useful for education and training of new surgical residents, and for demonstration and dissemination of novel surgical approaches around the world [25]. However, the absence of broadband communication infrastructure in remote rural communities or in military theatres, limits the ability to facilitate real-time video communication. While bandwidth limitations impose an upper bound on the bitrate of the videos to be encoded, a lower bound is imposed

by the low tolerance towards any compression-related artifacts on the diagnostically-important surgical content within the video. As a result, enabling real-time surgical telementoring, while balancing these seemingly conflicting requirements still remains a technically challenging problem.

2.1.1 Bandwidth requirements for telementoring

Several successful deployments of telementoring systems have been reported in literature. Demartines *et al.* attempted to assess the accuracy of telediagnosis in surgical cases by arranging weekly teleconferences between six university hospitals in four European countries over a period of two years [25]. The participants were presented with short edited clips of surgical procedures, and then asked questions such as "is the organ structure recognizable?", and " is the organ fine structure recognizable?". These teleconference were conducted using six Integrated Services Digital Network (ISDN) lines with a total channel capacity of 384 kbps. The authors reported that presentation of the surgical clips was rated positively by 72.2% of the participants, and the transmission bandwidth of 384 kbps performed satisfactorily, resulting in an accuracy of 93.3% in recognizing the organ structures, and an accuracy of 60% in recognizing the fine structures. However, the artifacts increased from 3% to 12% when the bandwidth was reduced to 256 kbps. Boanca *et al.* employed channels with average bandwidth between 800 kbps to 1024 kbps to compare the experience of surgeons in accessing the surgical field using a robotic camera, as opposed to a separate pan-zoom camera mounted on a tripod [10]. The surgeon was able to identify the surgical anatomy equally well in both cases. Schulam *et al.* developed a system for transmitting videos of laparoscopy procedures from the operative site to a central site (mentor location) using a single T1 line with a channel capacity of 1.54 Mbps [85]. These videos were captured with a frame-resolution of 176x144 (QCIF) and a frame-rate of 30 frames per second (fps), and were encoded using the H.261

video-coding standard [37]. All the seven cases were completed successfully and without any complications using their teleconsultation system. Several other examples of intra-hospital, local, and international telementoring systems were reviewed in [83] & [2]. In most clinically successful demonstrations of telementoring systems in general surgery, medium (384 kbps) to high (in excess of 1.5 Mbps) bandwidth resources were required for real-time delivery of fairly low-resolution medical videos. However, broadband communication infrastructure is nonexistent in remote rural communities and either absent, or damaged due to warfare in military combat zones. Bandwidth constraint has previously been cited as a rate-limiting resource that can render wireless, long-distance telementoring systems ineffectual [78]. As an example, Cubano *et al.* used "off-the-shelf" components to establish video, voice, and data communication between the USS Abraham Lincoln aircraft carrier cruising the Pacific ocean, and stations located in Maryland and California [23]. However, with the total bandwidth limited between 9.6 kbps to 28.8 kbps, video communication had to be reduced to very low frame-rates of 2-4 fps (typically videos are captured at the frame-rate of 24, 30 or 60 fps), and introduced delays of 2-20 seconds.

Clearly, the diagnostically-important surgical content within the videos place a lower bound on the bitrate of the encoded videos. Furthermore, procedures involving excessive dissection, suturing and knot-tying demand even higher communication capacity. Silva *et al.* performed 18 porcine pyeloplasty procedures using a robot and satellite links [24]. Their experiments concluded that telesurgery was not possible at a bandwidth below 4 Mbps. Hiatt *et al.* subjected controlled procedures of laparoscopic surgery to various levels of compression using the JPEG standard and displayed them to physician and nurse observers [42]. The authors concluded that clinically acceptable video compression for remote proctoring of laparoscopic procedures could be performed only at a bitrate of 1.5 Mbps (or higher). Similarly, Nouri *et al.* attempted to determine the necessary bitrate for transmitting medical-grade videos [71]. The

surgical videos were compressed using the MPEG-2 video-coding standard [68], and evaluated by a panel of experienced surgeons. The expert evaluations concluded that videos encoded with a bitrate of 3 Mbps (or higher) have a medically-satisfactory quality. A relatively new video codec (Windows Media Video 9) was used to enable mobile remote consultation for care of patients with acute stroke by Kim *et al.* [57]. Their study concluded that the remote specialists had a difficulty assessing patients with a stroke by viewing videos encoded with bitrates below 400 kbps.

All the above examples substantiate that medical-grade video is one of the most demanding data modality in terms of the necessary transmission bandwidth. Unfortunately, in remote settings, low-bandwidth wireless channels are the only possible mode of video communication. As per the feedback from our collaborators at the Telemedicine and Advanced Technical Research Center (TATRC) at Fort Gordon in Augusta, Georgia, very low bandwidth of around 128 to 200 kbps might be available for video communication with the forward military sites. To the best of our knowledge only one demonstration of surgical telementoring in low bandwidth scenarios has been reported in literature [36]. In most cases, lack of sufficient bandwidth tends to render clinically beneficial programs like remote surgical mentoring undeliverable. ROI video coding is a potential solution to address the constraints imposed by limited bandwidth channels. By disproportionately allocating larger information rate resources to the ROI as compared to the background (BG), it could be possible to encode the surgical field with a visual quality that is appropriate for medical assessment.

2.1.2 Region of interest video coding

Several methods have been proposed for introducing ROI support in encoded videos. Wong *et al.* proposed a ROI-based, channel-adaptive, source-coding scheme for wireless video transmission [112]. Channel bandwidth was computed based on the channel

state information, and the ROI and the BG were then adaptively encoded with different and varying compression ratios determined by the channel bandwidth. The BG was dropped altogether when necessary. Chai *et al.* proposed and implemented two ROI-coding strategies [15]. Their maximum-bit-transfer (MBT) strategy assigned the highest permitted compression level to the BG, and the lowest possible compression level to the ROI in a manner that ensured that the target bitrate was not exceeded. Their joint bit-allocation strategy allocated bits to the ROI and the BG based on the size, motion, and priority of each region. They also implemented a user-defined mode, where the user could adjust bit consumption using a scale that ranges from "not-coding-foreground" to "coding-foreground-only". A central ROI was assigned the highest priority, and multiple regions around the ROI were assigned decreasing priorities before reaching the background in [1]. The background was assigned the lowest priority. To introduce a graceful degradation in the picture quality going from the ROI to the BG, higher priority regions were allocated more bits as compared to the lower priority regions. This method specifically targeted a two-layer approach for rate control, and assigned bits at the group-of-pictures (GOP) and the frame levels. Given the iterative nature of the rate-allocation algorithm presented in [15], and the complexity of the algorithm presented in [1], a real-time implementation of these methods seems infeasible. Although real-time ROI coding is targeted in [112], the results were demonstrated on a five fps source video, which would be considered too low for surgical telementoring applications.

The algorithms presented in [15] and [112] were implemented using the H.261 and the H.263 [38] video-coding standards respectively, while the algorithm in [1] was implemented by modifying the H.264 reference software. H.264 is currently the most popular video-coding standard. As a result, several approaches have been proposed to introduce ROI in an H.264 compliant fashion. Such techniques primarily rely on using the Flexible Macroblock Ordering (FMO) and the Arbitrary Slice ordering

(ASO) tools available in the baseline and the extended profiles of the H.264 standard. Using these tools, it is possible to divide a frame into multiple slice groups. Each slice group is composed of several independently decodable slices. Every macroblock of the picture can be freely assigned to one of these slice groups. Thus for example, one slice group can represent all the macroblocks of the ROI, and another slice group can be used represent all the macroblocks of the BG. In [26], FMO of Type 2 was used to divide the picture into a number of regions of interest and one region of disinterest. Further, the regions of interest were compressed at a higher bitrate than the region of disinterest. The region of disinterest was entirely dropped in the extreme case. Leuven *et al.* proposed an user-interactive implementation in which the user-defined regions of interest were signaled using a FMO of Type 2, Type 3 and Type 6 [100]. Further, a concept of an extensible ROI (x-ROI) was introduced to smooth out the transition between the ROI and the BG. Once the shape and the size of the ROI were determined, each picture was compressed using predetermined quantization parameters (QPs). For example, the set {25-30-35} would imply a QP of 25 for the ROI, a QP of 30 for the x-ROI and a QP of 35 for the BG. In [5], the ROI was signaled using a standard-compliant FMO mapping function and was combined with a QP-adaptation scheme to control the source rate. However, due to coding inefficiencies and additional decoder complexity associated with FMO, most H.264 encoders do not implement these error-resiliency features. FMO is also excluded from the most commonly used "main" and "high" profiles of the H.264/AVC standard. Furthermore, most of the techniques described so far, are blind to the medical content of the videos, and hence untested in their ability to support telementoring applications. For example, the schemes proposed in [112] & [26], BG was dropped entirely in extreme cases. However, complete loss of background context can be distracting for the medical experts examining the video (see Section 2.3 & [81]). The next sub-section reviews a few techniques which are targeted specifically for ROI-compression

of medical videos.

2.1.3 ROI coding of medical videos

Medical videos are evaluated by experts who require the videos to be diagnostically lossless. However, most ROI-coding algorithms proposed in literature are not strictly targeted towards medical applications. Very few approaches have specifically addressed ROI coding of medical videos. Yu *et al.* did not explicitly introduce the notion of a ROI, but presented a new measure to estimate motion complexity, and also proposed a perceptual bit-allocation method to heavily compress the non-medical content within a medical sequence [114]. For example, the authors suggested compressing the black background of an "echocardiography" sequence more heavily, as compared to the cardiac imagery in the picture. Although the PSNR and bitrate gains achieved by ROI encoding were negligible, the subjective evaluations indicated an improvement in the perceptual quality [114]. Perceptual quality however, is not the same as medically or diagnostically-acceptable quality. DL of encoded videos can only be determined by medical experts. However, the subjective evaluations in [114] involved a single "medical doctor". In [90], the authors developed a "gaze map" indicating the ROI in the frame by tracking the eye movements of the surgeon. Using this gaze map, the videos were pre-processed with a filter to aggressively remove texture information from regions outside the ROI, giving as much as 50% reduction in the bit-rate of the encoded video. However, the impact of removing the so-called "redundant information" on the complexity of the encoder, and the diagnostic quality of the videos were not evaluated by experts. In [97] the ROI in carotid ultrasound videos was identified using a visual attention model and signaled using a saliency map. Once the ROI was identified, the BG was blurred and the video was compressed using MPEG-2 and MPEG-4 [69]. Since the BG needs to be blurred to reduce its quality in the encoded video, the encoded quality of the ROI has to be varied indirectly by

varying the strength of the median filter. Although a gain in the coding efficiency was reported in [97], the authors also noted that the medical experts evaluating these videos did not express a significant preference towards the ROI-encoded videos. An elastic ROI was introduced in telepediatrics videos in [80, 81]. Expert feedback from physicians was used to determine the bitrate at which the mathematically lossless (ML) and DL quality was achieved when encoding various pediatrics videos. This information was then incorporated in the encoder design by using a state machine to determine the bitrate that should be allocated to the ROI and the BG of telepediatric videos. Based on subjective evaluations by physicians, their study concluded that ROI-encoded videos can be deemed diagnostically lossless at bitrates around 500 kbps as compared to 1077 kbps for the uniformly encoded videos. However, as will be seen in Section 2.3, considerable amount of inter-expert variability and content dependancy may exist when evaluating the quality of surgical sequences. This makes it impossible to allocate a single bitrate [81], or a single QP value [100] for the ROI across different surgical sequences, while ensuring a DL quality. Still, based on the existing literature, it can be asserted that ROI encoding can prove valuable for compression and transmission of surgical videos in low bandwidth scenarios. In the next few sections, we will propose a method to introduce a flexible, interactive and low-complexity ROI support in a standard-compliant fashion for enabling remote mentoring of surgical procedures.

2.2 Region of interest compression for video delivery in low bandwidth scenarios

As seen in the previous section, encoding surgical videos with a ROI support can improve the quality of the surgical field delivered over very-low bandwidth channels. Besides the challenge of delivering a real-time, diagnostically lossless video over bandwidth constrained channels, several other restrictions may be imposed by a telementoring application. The operating room located in a rural surgical facility or

military field hospital may have limited computing capabilities, and may not be able to support high complexity algorithms for encoding and transmitting the surgical videos. Furthermore, to support multiple, and possibly mobile decoding platforms at the mentor site, the video needs to be encoded in a standard-compliant fashion. Given the above requirements, this work focuses on adapting an existing H.264 encoder to introduce flexible, interactive, and diagnostically lossless ROI in surgical videos, as described below.

2.2.1 Objectives

The desired outcome of this part of the research is a military surgical paradigm where less experienced surgeons deployed at forward sites can be mentored by more experienced specialists from a rearward remote site, enhancing in-theater surgical options by bringing scarce expertise into play. Given the surgical nature of the videos, the first and foremost objective of the presented research is to determine the effectiveness of establishing a ROI in encoding a diagnostically lossless surgical field. The second aim of the proposed research is to develop and evaluate a prototype demonstrating a surgical telementoring application.

2.2.2 Equipment

For this project we used a head-mounted camera configuration to capture the surgical procedures. In consultation with physicians at the Eisenhower Army Medical Center (EMAC) at Fort Gordon, GA, we determined to use to the Luxtec surgical video system [64]. As seen in Figure 4, this system consists of a head-mounted camera and a light source mounted on the same headset as the camera. In the video system, the camera output is fed to a MPEG-2 encoder, and optionally to a liquid crystal display (LCD), and a digital video recorder (DVR) for recording the captured video. The camera output has a spatial resolution of 720x480 pixels and a frame rate of 30 fps. Ideally, we would have preferred gaining access to the uncompressed video

output of the system. Unfortunately, due to the relatively closed (black-box) nature of the surgical system, accessing uncompressed video output from the camera is not very feasible. However, the built-in MPEG-2 encoder compresses the captured videos at a very high bitrate of around 6.4 Mbps. Based on the manufacturers testing, and existing literature, a bitrate higher than 6 Mbps can be safely considered to be diagnostically and visually lossless. As a result, we simply decoded these MPEG-2 compressed videos, and treated them as raw or uncompressed videos for the purpose of our experiments.

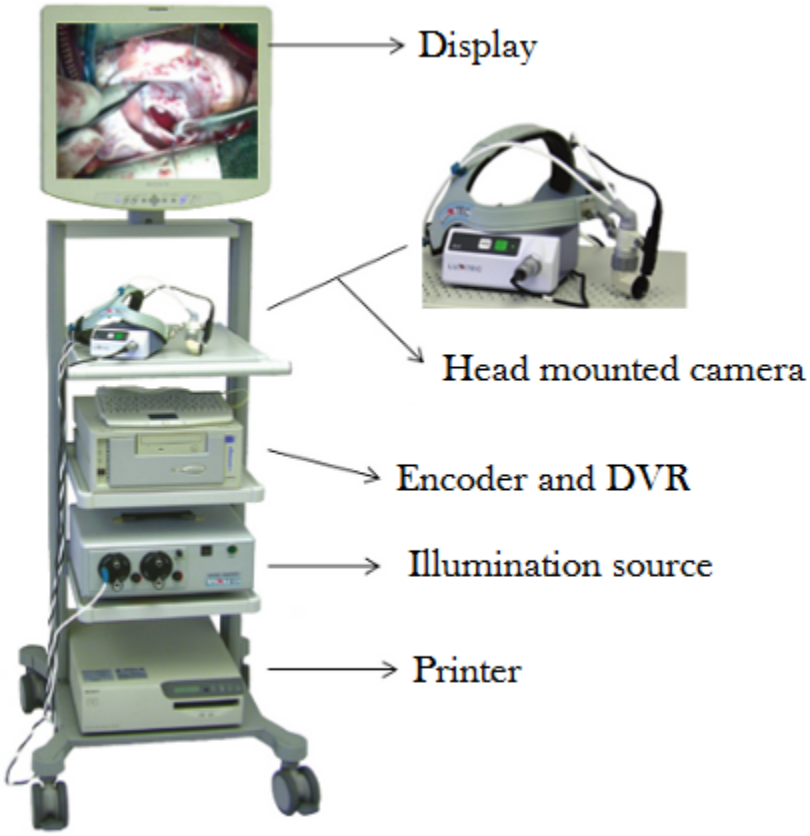
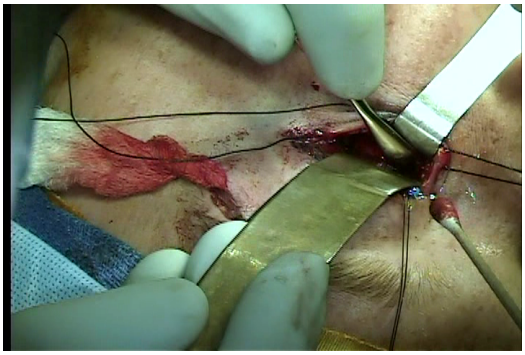


Figure 4: Luxtec surgical video system [64].

(a) Transconjunctival incision

(b) Left lateral skin closure

(c) Intraoral vestibular access



(d) Left orbital rim fracture

Figure 5: Example clips in the surgical video database

2.2.3 Database

Multiple video clips of maxillofacial surgery were recorded by a surgeon using the head-mounted camera. The clips included a variety of surgical procedures such as incision, suturing, tooth-extraction etc. Figure 5 shows some of the clips in the database.

2.2.4 Region of interest video compression

The design of the proposed system is based on the following assumptions,

1. The upper bound on the channel capacity for video transmission is 200 kbps.
2. One-way video communication capability is available to enable the remote mentor to view the surgical procedure.

3. Two-way audio communication capability is available to enable the remote mentor to interactively engage the surgeon during the procedure.

Given the design requirements of facilitating live video communication, we choose the x264 software for encoding [67]. X264 is a free and open source library and software for real-time encoding of video streams in the H.264/MPEG-4 AVC format. An overview of the various rate-allocation modes available in x264 can be found in [67]. We use the average bitrate, 1-pass mode to achieve the target bitrate in real time. Without any modifications, x264 can compress a 720x480p, 30 fps input video with a processing speed well over 30 fps using a generic desktop computer. This processing speed easily meets the real-time requirement of our system.

To introduce a flexible and interactive ROI in the picture, we modify the quantization coefficients of each macroblock of each frame. By differentially quantizing different regions of the video frame we can create a low bitrate background region, and a high bitrate ROI. This strategy introduces ROI encoding in a standard-compliant fashion, thus ensuring that any decoder capable of decoding H.264 bitstreams will also be able to decode the ROI-encoded videos. The H.264 standard specifies 52 values (0-51) for the QP, and each value of the QP maps to a certain step size of the H.264 quantizer. As the value of QP increases, the quantizer step size also increases. Furthermore, the step size doubles for ever increment of six in the QP value. In the x264 source code, we identified an entry point to insert custom macroblock quantizer coefficient offsets to implement the ROI encoding capability. ROI is implemented in the x264 encoder by supplying a positive or negative 'QP-offset' value to each macroblock of the frame. Thus the ROI can be established by supplying a negative QP-offset value to reduce the QP of the ROI macroblocks and simultaneously supplying a positive QP-offset value to increase the QP for the BG macroblocks. X264 adaptive quantization in the 1-pass mode assigns QP values by estimating various parameters such as the motion and texture complexities within the picture. The

user-defined QP-offsets are applied on top of the QP decisions made by the x264 rate-allocation algorithm. In [100], predetermined QP values are forced on the ROI and the BG regions. This has the disadvantage of indiscriminately quantizing the regions of varying complexities within the ROI and the BG with the same QP. In contrast to the approach in [100], by applying user-defined QP-offsets on top of the decisions made by x264, we are able to preserve the rate-allocation decisions made by x264 while still allocating larger bits to the ROI as compared to the BG. However, since the QP-offsets are applied on top of the QP decisions made by x264, we need to ensure that the target bitrate is not affected by the externally induced changes in the quantization parameter. After running multiple experiments on the test videos, we empirically determined to restrict the QP-offset values for the ROI between '0 to -10' and the QP-offset values for the BG between '0 to $+8$'. To simplify the application of QP-offsets, a QP-offset control was provided to the user. As the QP-offset control varied from 0 to 10, the ROI visibility varied from no-ROI (or uniformly compressed frame) to a strong ROI (and a heavily compressed background). Figure 6 shows the effect of three different QP-offset settings: 0, 6, 10 on the *left lateral skin closure* sequence. Qualitatively, these settings correspond to uniformly compressed, or no-ROI (0), medium-ROI (6) and strong-ROI (10). As seen from Figure 6, the visibility of the suture clearly improves with a strengthening ROI.

Given the head-mounted nature of the camera, the surgical field could move over the entire frame during the procedure. As a result, the ROI encoded in the transmitted video would also have to follow this movement from frame-to-frame. This problem can be partially simplified by assuming that the remote mentor can request the surgeon to momentarily hold his or her head still. Given this simplification, a low-complexity and flexible ROI was introduced in the system. This flexibility can be achieved by dividing the frame into a 3x3 grid (or 9 cells), such that we have cells (1,2,3) on Row 1, cells (4,5,6) on Row 2, and cells (7,8,9) on Row 3 as seen in Figure

(a) QP-offset=0

(b) QP-offset=6

(c) QP-offset=10

Figure 6: Effect of changing QP-offset value on the ROI and the BG quality of the *left lateral skin closure* sequence.

7. Now we have the flexibility to choose five different locations of ROI, each of size 2x2 (or four cells). These are: (a) Northwest (1, 2, 4 and 5), (b) Northeast (2, 3, 5 and 6), (c) Southwest (4, 5, 7 and 8), (d) Southeast (5, 6, 8 and 9) and (e) Center. The center ROI region is composed of the central cell (5) and extends half-way into each of its neighbors (1-9). Figure 8 shows the various possible ROI locations in our implementation.

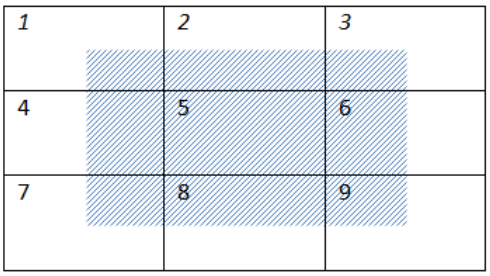


Figure 7: Illustration of the nine cells and the central-ROI.

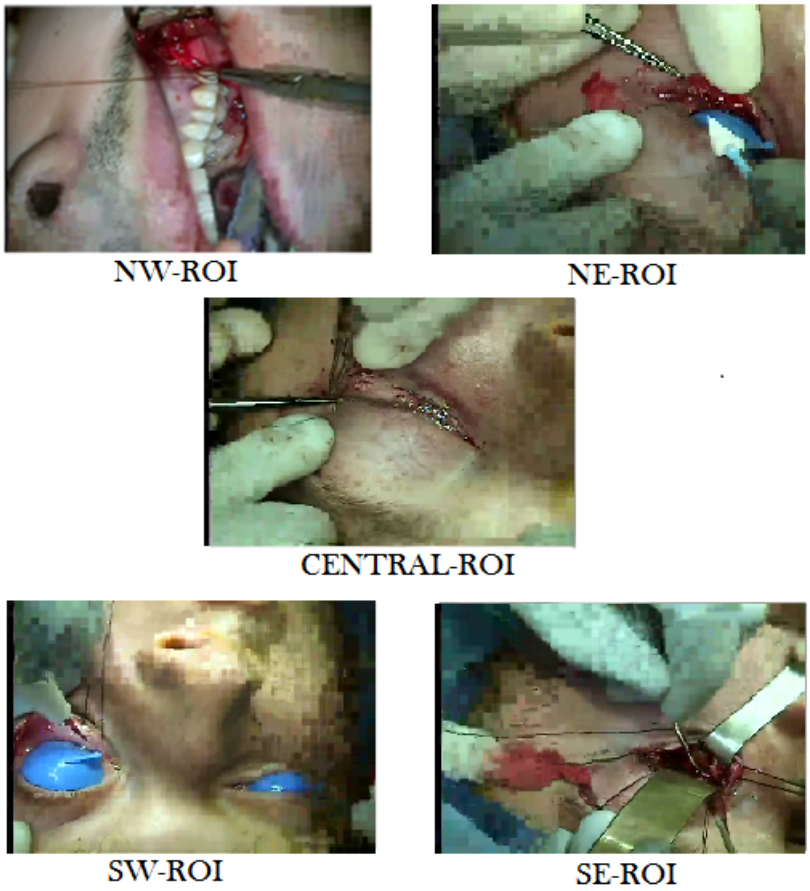


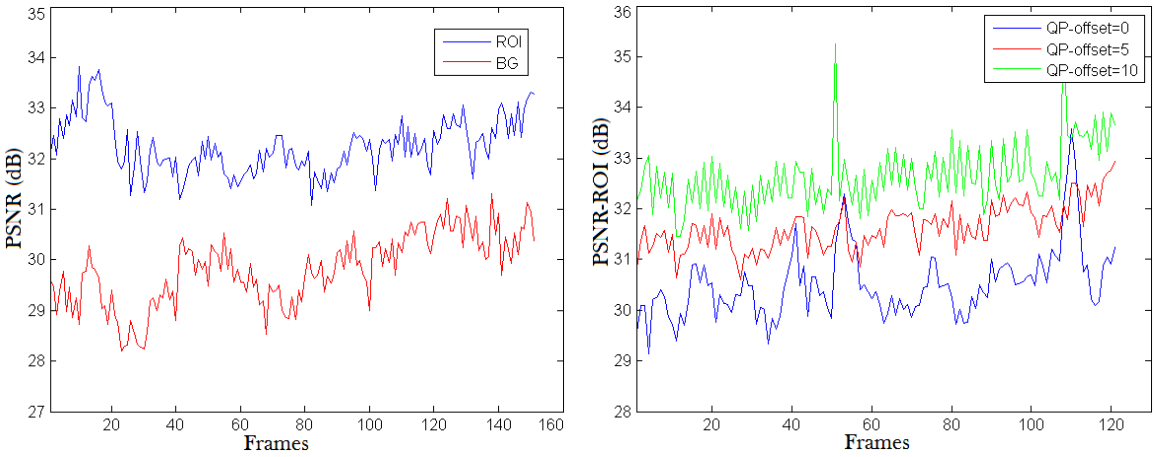
Figure 8: Various possible locations of the ROI in the proposed system.

Some ROI encoding approaches implement automatic detection and tracking of the ROI using various gaze tracking [90], or segmentation [63] algorithms. Introducing such complexities for a surgical telementoring application is not only unnecessary, but almost undesirable. It is important to note, that the ultimate aim of the surgical telementoring system is to deliver a DL-quality surgical video to the mentor. Given the level of inter-expert subjectivity involved in evaluating surgical videos, the best individual (or algorithm) to determine the location and the quality of the ROI is the mentor actually making the remote diagnosis. Hence, instead of transmitting videos which are encoded using a ROI defined by a person, or an algorithm at the encoder (operative site), an interactive user-interface is provided to the remote mentor. This enables the mentor to view the live video, and signal a change in the quality and the

location of the ROI as and when necessary.

2.3 Results and Discussion

Figure 9a shows the PSNR of the ROI and the BG for the *intraoral vestibular access* sequence compressed at 150 kbps with a QP-offset value of 10 (strong ROI). As seen from the figure, the PSNR of the ROI is boosted at the expense of the PSNR of the BG. Figure 9b shows the variation of the PSNR of the ROI with varying values of QP-offset. Clearly, as the value of QP-offset increases, the PSNR of the ROI also increases.



(a) PSNR of the ROI and the BG with QP-offset=10

(b) PSNR of the ROI with different QP-offsets

Figure 9: Analysis of the *intraoral vestibular access* sequence.

Figure 10 illustrates the ability of the proposed ROI-based encoder in increasing the apparent bandwidth of the transmission channel. Although the video in Figure 10a, and the video in Figure 10c are encoded at the same bitrate (156 kbps), the ROI of the video (Northwest) in Figure 10c appears to have a better perceptual quality. This difference is particularly noticeable around the incision and near the tip of the surgical instrument. Moreover, the ROI of the video in Figure 10c which is encoded at 156 kbps appears to have the same visual quality as that of Figure 10b which is encoded at 256 kbps. Thus, for video transmission over a channel whose bandwidth

(a) 156 kbps, without a ROI

(b) 256 kbps, without a ROI

(c) 156 kbps, with a ROI

Figure 10: Frame of a surgical video compressed at (a) 156 kbps without ROI, (b) 256 kbps without ROI, and (c) 156 kbps with ROI.

is limited to only 156 kbps, the ROI-encoded video (Figure 10c) seems to be a better choice compared to the uniformly-compressed video (Figure 10a).

Although ROI encoding gives a PSNR gain in the ROI of the picture (Figure 9a), and appears to have a higher perceptual quality (Figure 10), the diagnostic quality of the ROI-encoded videos can only be evaluated by experts, or in this study, by surgeons. To establish the DL criterion, and also to evaluate the usefulness of ROI encoding in surgical videos, subjective evaluation involving multiple surgeons were conducted. The test design is briefly described below:

1. Three different surgical videos (Figure 5a, Figure 5b and Figure 5d) were chosen for the test. Each video was compressed at three different bitrates: 125 kbps, 150 kbps and 200 kbps and three different QP-offset values: 0, 6 and 10.

2. The entire evaluation was presented in the form of nine "test sessions". Each test session consisted of three videos which were all encoded at the same bitrate, but with different QP-offsets.

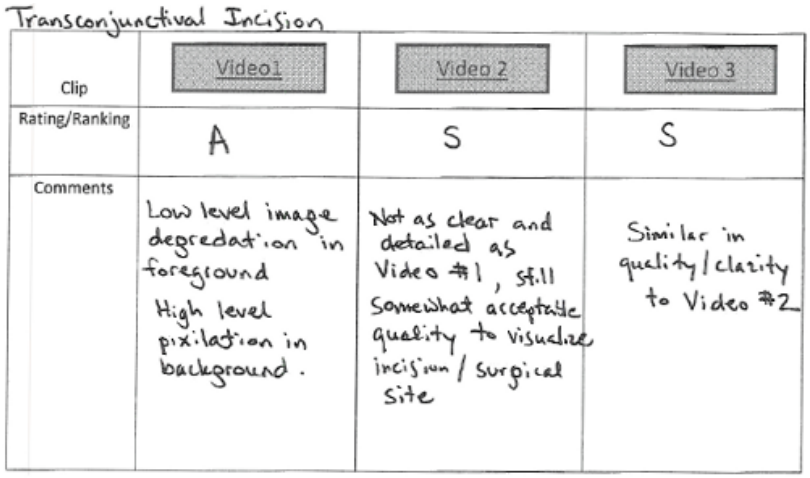
3. The evaluators were instructed to view each video and give a rating of acceptable (A), somewhat acceptable (S) or unacceptable (U) to each video. The evaluators were also encouraged to fill out the comments section with qualitative observations.

4. Since the evaluators were unfamiliar with video-coding and region-of-interest technology, a "training session" was included before the actual evaluation. The training session comprised of videos demonstrating all possible "effects" that would appear in the evaluation videos. Thus, the training videos consisted of surgical videos encoded with varying bit-rates (and hence varying levels of compression artifacts such as blockiness), and varying strengths of the ROI. The training videos were not included in the test sessions.

5. To avoid biasing the responses of the evaluators by presenting videos in a predictable pattern, the display order of all the videos for different test sessions was completely randomized. This made it impossible for the evaluators to determine the bitrate and QP-offset value of the video based on the order in which it appeared in the evaluation.

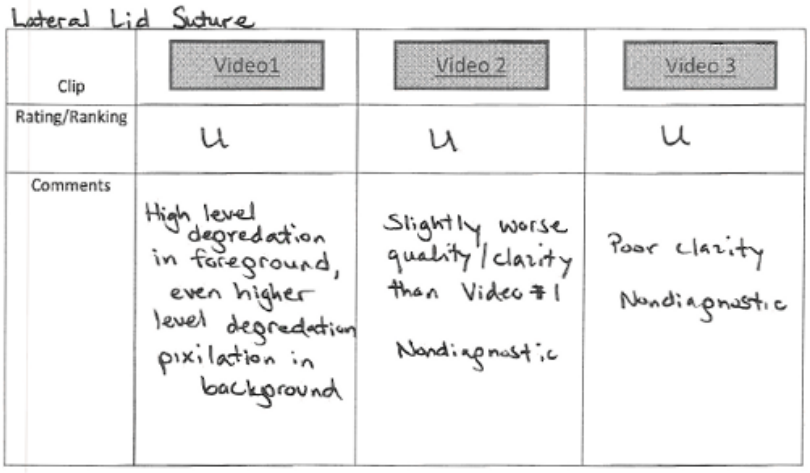
Figure 11a shows an example response of one of the evaluators for Test Session 2. The three videos displayed in this test session were all compressed at 200 kbps and with QP-offset of '10' for Video1, '6' for Video2 and '0' for Video3. As seen from the figure, all the videos were rated as 'A' or 'S'. Based on the ratings and comments given by a surgeon for this test session (Figure 11a), there seems to be a preference towards a stronger-ROI over the medium or no-ROI videos. However, as seen in Figure 11b, in contrast to Test Session 2, all the videos in Test Session 4 were rated 'U'. All the

videos in Test Session 4 were compressed at 125 kbps. Thus the response of the same surgeon for Test Session 4 indicates that no amount of ROI coding is sufficient to achieve diagnostic quality for this video encoded at 125 kbps.


Transconjunctival Incision

Clip	Video1	Video 2	Video 3
Rating/Ranking	A	S	S
Comments	Low level image degredation in foreground High level pixilation in background.	Not as clear and detailed as Video #1, still somewhat acceptable quality to visualize incision / surgical site	Similar in quality/clarity to Video #2

(a)


Lateral Lid Suture

Clip	Video1	Video 2	Video 3
Rating/Ranking	U	U	U
Comments	High level degredation in foreground, even higher level degredation pixilation in background	Slightly worse quality/clarity than Video #1 Nondiagnostic	Poor clarity Nondiagnostic

(b)

Figure 11: Evaluation of (a) Test Session 2 and (b) Test Session 4 by the same surgeon.

A total of eight surgeons and seven dental residents evaluated the ROI-encoded videos. Table 1 shows the aggregated evaluations of all participants. The clips were displayed to the participants in the order seen in Table 1. The final three columns of the table indicate the number of times a video was rated either 'A', 'S' or 'U'. These evaluations were further aggregated as seen in Table 2, to determine the effect

of varying bitrate and varying quality of the ROI using QP-offset on the quality of the encoded videos.

Table 1: Aggregate of evaluations received from all participants

STIMULUS	BITRATE	OFFSET	#A	#S	#U
CLIP2	125k	Offset_10	3	8	4
		Offset_06	8	3	4
		Offset_00	3	4	8
CLIP1	200k	Offset_10	8	5	2
		Offset_06	10	5	0
		Offset_00	10	4	1
CLIP2	200k	Offset_10	5	7	3
		Offset_00	8	6	1
		Offset_06	8	6	1
CLIP3	125k	Offset_10	2	7	6
		Offset_00	4	6	5
		Offset_06	6	6	3
CLIP3	200k	Offset_06	8	4	3
		Offset_00	7	7	1
		Offset_10	9	3	3
CLIP1	150k	Offset_06	7	5	3
		Offset_10	6	8	1
		Offset_00	9	6	0
CLIP3	150k	Offset_10	4	8	3
		Offset_06	5	9	1
		Offset_00	8	2	5
CLIP2	150k	Offset_06	6	8	1
		Offset_00	2	7	6
		Offset_10	7	7	1
CLIP1	125k	Offset_06	9	4	2
		Offset_00	8	5	2
		Offset_10	9	3	3

From Table 2, two following two trends can be spotted:

1. Higher bitrate is always preferred. Thus, the evaluators always rated a video compressed at 200 kbps as possessing better or the same quality as that of a video compressed at 125 kbps or 150 kbps. Furthermore the difference in the number of 'A' and 'U' ratings for a video compressed at 200 kbps is much higher

Table 2: Summary of evaluations by all participants

		A	S	U	#A - #U
Bitrate	125	52	46	37	15
	150	54	60	21	33
	200	73	47	15	58
QP-offsets	0	60	45	30	30
	6	67	50	18	49
	10	52	58	25	27

than for the other two encoding bitrates under consideration.

2. It was also observed that although ROI-encoded videos are preferred over uniformly compressed videos, the strongest ROI quality was not always preferred. As seen in Table 1 there were multiple instances where evaluators rated video encoded with QP-offset of 10 to be lower than video encoded with QP-offset of 6. This observation was also confirmed in Table 2, where the difference in the number of 'A' and 'U' ratings for QP-offset of 6 is significantly larger than that for videos with QP-offset of 0 or 10. Although sufficient qualitative comments were not available to explain this observation, as seen in Figure 12, some reasons such as "weird depth issue" for the videos compressed with strong-ROI were noted. Videos in Figure 12 were all compressed at 200 kbps, but with QP-offsets of 6, 0, and 10. This indicates that although increasing the ROI-quality at the expense of the BG can help, boosting the quality of the ROI to the point where the BG is highly distorted can be distracting for the expert evaluators. This observation is consistent with the study in [81], where the authors have reported that background context, or overview information is important to the physicians, and have also noted that, "DL quality in the ROI is necessary and is the most important issue, but is itself not a sufficient condition for the overall assessment of the patient."

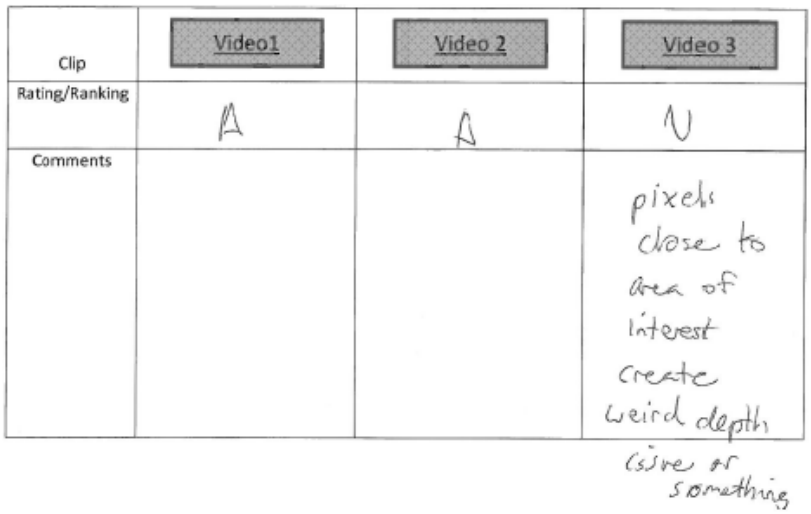


Figure 12: One evaluation of Test session 5.

2.4 ROI encoding prototype

A ROI encoding prototype was implemented to demonstrate real-time ROI encoding and streaming of surgical videos. This implementation utilized the open source VLC video player and streaming server [73], and the open source x264 encoder. A basic user interface (UI) which runs on a web browser was also implemented to allow the remote mentor to control the location and the quality of the ROI within the frame. Figure 13 shows a screen-capture of the prototype in action. The prototype is enabled by launching the VLC player in a streaming server mode of operation. Encoder parameters such as single-pass encoding and small look-ahead buffering are chosen to facilitate low-latency and real-time encoding of the surgical video, and the target bitrate is set equal to the actual capacity of the channel. The encoded video is streamed over the network using Hypertext Transfer Protocol (HTTP). The VLC encoder periodically checks the contents of a local file which contains information regarding the current location of the ROI (NW, NE, Center, SW, SE), and also the current value of the QP-offset (0-10) requested by the remote mentor. The client UI is a PHP: Hypertext Preprocessor (PHP) script running on the web server which allows the client to change the location and the strength of the ROI remotely, and in

real time.

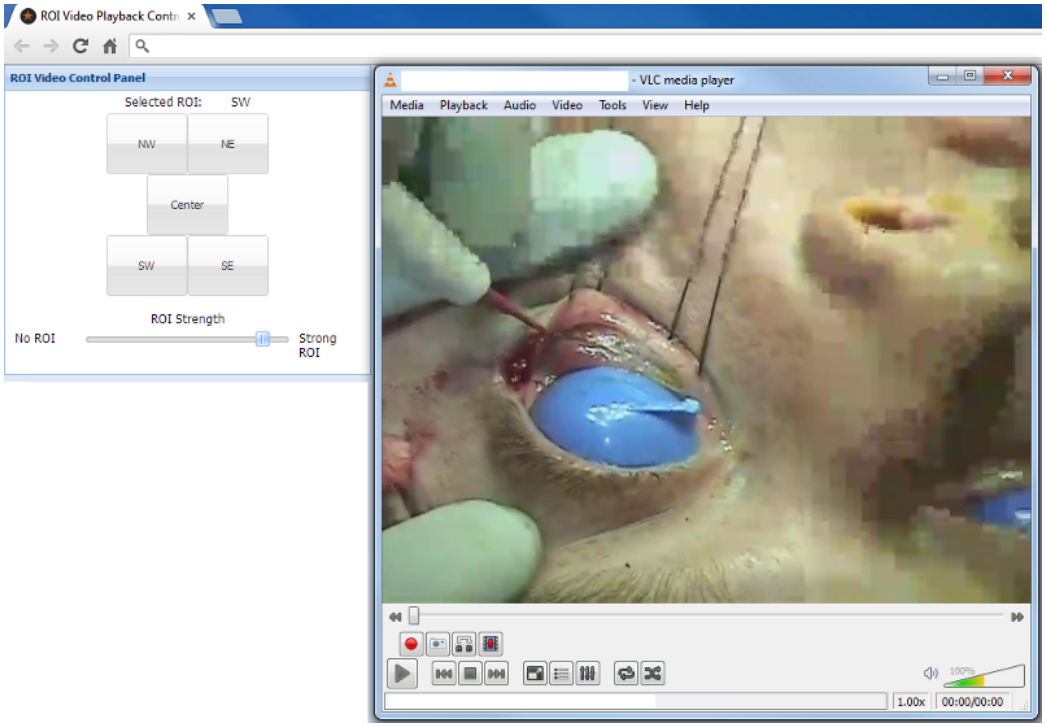


Figure 13: The client UI of the ROI evaluation prototype.

The ROI prototype was also evaluated by three experts. The users were given the opportunity to experiment with the prototype for as long as they wished, and then asked to provide qualitative answers to four questions structured to prompt feedback from the users regarding the prototype, as well as the ROI-coding technology itself. The questions and a few select responses are enumerated below,

1. **Is it useful to have buttons to control the location of the ROI? Is this location control easy to use?**

   "It is very useful to have buttons to control ROI location." "It is important to allow the surgeon to have the option to change the location." "It will be nice to have it (the ROI) outlined or highlighted for a second when it starts."

2. **Is it useful to have a slider to control the strength of the ROI? Is this strength control easy to use?** " It was very useful to have a slider control

the strength of the ROI. If bandwidth is a concern the strength control can be adjusted for best clarity available at the time."

3. **What additional features would you like to see to improve the usefulness and the usability of the tool?** " include adjustment options for - color, sharpness, brightness, etc. to improve contrast between anatomic structures," "Maybe (include) some numbers impacting quality of video like kb/s or resolution." " .. 10 grades of ROI level are too much and not very convincing. None - Some - Lots (1-5-10), however, may be useful in some situations. In general, on the other hand, having it on or off may be all you need."

4. **After having used the prototype, what is your opinion on the usefulness of ROI for viewing remote videos?** "ROI for viewing remote videos can be a very beneficial option in austere environments. ..allowing an opportunity to make critical decisions and provide insight that would positively impact morbidity."

2.5 Non-ROI methods for surgical telementoring over low-bandwidth channels

As detailed in Section 2.1, the challenges involved in transmitting high-quality surgical videos over low-bandwidth channels are immense. However, ROI coding may not be the only solution to address these challenges. Martini *et al.* designed a network-aware video compression and transmission system to address the contrasting requirements of low compression ratio, and high loss protection for transmission of medical videos over low bandwidth channels [66]. Their system varies parameters such as frame rate, QP values for the I and P-frames, and the GOP-length of MPEG-4 video, in response to the received video quality measured using metrics such as PSNR. However, it is well known that PSNR does not relate well with the visual or the diagnostic quality of the decoded video. In [18], a mobile teletrauma system was

implemented over a commercial 3G cellular connection whose bandwidth fluctuated between 153 kbps and 60 kbps. Congestion on the channel during low data-rate intervals was managed by varying the frame-rate of the transmitted medical video. Although the transmitted videos used in their work had a fixed frame resolution of 320x240, the frame rate varied between 2-25 fps. The videos were encoded using the motion JPEG (M-JPEG) standard to give a compression ratio between 15:1 to 25:1. Again, the frame-rate adaptation in this system was done entirely by estimating the link bandwidth, and did not consider the implications of varying the frame-rate on the diagnostic quality of the video sequence received by the remote consultant. In [27], the patient and network status was monitored to determine the quality of transmitted video. Patient videos were encoded using the scalable video coding extension of H.264 [86]. Only the base layer was transmitted during normal patient status, but a high quality stream with a base layer of 128 kbps, and two enhancement layers, each encoded at 256 kbps were transmitted during urgent patient status. PSNR and Structural Similarity Index (SSIM) were measured to demonstrate that the received streams have an "acceptable quality". Unfortunately, PSNR and SSIM have no fixed relation to the diagnostic quality of medical videos. H.264/SVC was also employed in [76] to encode and transmit ultrasound videos at VGA and QVGA resolutions, and frame rates of 10, 15, 20 and 30 fps over wireless and cellular networks. Their study concluded that 30 fps VGA sequences could be transmitted over the 802.11g (wireless) connection, but only 10 or 7 fps QVGA sequences could be transmitted over the 3G cellular connection while still retaining most of the diagnostically important information in the video. Using H.264/SVC, the base layer was encoded at QVGA layer, while an additional enhancement layer permitted transmission of sequences at VGA resolution, provided sufficient bandwidth was available. Scalable coding tends to increase the encoder and decoder complexity, and hence is usually not provisioned

in most devices. However, all the work reviewed in this section indicates that DL-quality transmission of medical videos over bandwidth limited channels is possible by adaptively scaling the frame rate and/or frame resolution.

Since the existing work presented above does not deal with specific modality of surgical videos, a small controlled test was conducted to determine the effectiveness of scaling the frame rate and frame resolution in enabling remote surgical mentoring. Three surgical videos with varying amount of spatial and temporal complexity were chosen from the database, and then scaled to obtain four different combinations of frame rates and resolutions:

(a) 720x480, 30fps (original)

(b) 720x480, 15 fps (temporally downsampled)

(c) 360x240, 30 fps (spatially downsampled)

(d) 360x240, 15 fps (temporally and spatially downsampled)

These videos were further encoded at the three different bitrates: 125 kbps, 156 kbps and 200 kbps using the H.264/AVC standard. These downsampled and encoded videos were then evaluated by three experts using a similar setup as the one described in Section 2.3. Following observations were made from this experiment:

1. No improvements in the ratings were noted for the videos in case (b). This indicates that dropping the frame rate down to 15 fps (while maintaining the same bitrate) does not degrade or improve the diagnostic quality of a sequence from 'U' to 'S', or 'A'.

2. Overall rating for multiple sequences improved from either 'U' or 'S' to 'A' when the frame resolution was down-scaled in cases (c) and (d). One video used in this test was common to evaluations described in Section 2.3. For this video,

ROI support had also facilitated similar improvement in the rating of the clip from 'S' to 'A'.

While this test is too small to make a compelling case in favor of dropping frame rates or resolutions, it does seem to indicate that insufficient bandwidth, not reduced frame rate and frame resolution is the barrier to achieving DL quality in compressed surgical videos. More comprehensive testing in the future could validate that both: ROI and SVC technologies can potentially deliver surgical video with DL-quality over low-bandwidth channels.

2.6 Summary

This chapter presented a method to introduce ROI support in live patient video for enabling remote mentoring of surgical procedures over low-bandwidth channels. The ROI was implemented in a low-complexity and H.264-compliant fashion. Further, diagnostically lossless delivery of the surgical field was facilitated by enabling the remote mentor to interactively choose the location and the strength of the ROI. Objective and subjective evaluations indicated that ROI-encoded videos can increase the apparent bandwidth of the low-bandwidth channels, and facilitate surgical telementoring. ROI-based video coding was thus included as one of the video coding schemes in the ASVCD suite of solutions. Further, a prototype employing ROI video coding for enabling surgical telementoring was developed, and received favorable evaluations by experts. The findings from this Chapter were published in [53], and are being considered for commercialization in the future.

CHAPTER III

MULTIPLE REPRESENTATION CODING OF STREAMING VIDEO FOR GRACEFUL RECOVERY FROM SIGNAL LOSSES

3.1 Introduction

The ease in creating and publishing video content afforded by the current generation of cheap and powerful mobile multimedia devices has rapidly increased the ability and the desire of consumers to generate and share high quality video content. In 2011, Google reported that 48% of smartphone users watch video on their smartphones [33]. As the number of smartphone users continues to rise, the amount of wireless and cellular video access is also expected to grow rapidly. Not surprisingly, video is already one of the most dominant sources of traffic over the Internet. According to a forecast by Cisco Systems, video data is expected to comprise 86% of the total traffic over the Internet by 2016 [91]. Unfortunately, in the mobile multimedia communication paradigm, much of this video data is distributed over channels prone to jitter, delay and bandwidth fluctuations, and abundant packet losses. Video communication can be severely hampered due to presence of such impairments on the channel. For example, a mobile video-streaming session can be interrupted due to the loss of signal experienced when riding an elevator, or traveling aboard an underground railway. As illustrated in Chapter 1, loss of signal eventually translates into loss of multiple frames and seconds of the video sequence, and severely degrades the quality of experience delivered to the end-user. The end-user in the case of telemedicine applications studied in Chapter 2, is a remote surgical mentor, expecting video-communication at a diagnostically-lossless quality. For more generic video streaming applications,

the end-user is a viewer initiating a mobile or wireless video access, and expecting a consistently high-quality video. Unfortunately, supporting such high-quality video services over unfriendly environments prone to burst losses and signal loss intervals remains a huge technological challenge.

Due to its pervasiveness, the problem of error-resilient video delivery has been studied for several decades. Numerous error-resiliency tools for video transmission have been proposed for the H.26x family of video-coding standards [59, 107]. In H.263+, up to 12 optional error-resiliency modes such as FEC, reference picture selection (RPS), and modes that facilitate organization of the picture into independently decodable slices and segments have been specified. Error-resiliency tools such as FMO and ASO have also been defined in the baseline and extended profiles of the H.264/AVC video-coding standard. Using these tools, it is possible to divide a frame into multiple slice groups. Each slice group is composed of several independently decodable slices. Every macroblock of a picture can be freely assigned to one of these slice groups. By grouping non-adjacent macroblocks in a single slice group, FMO can scatter the burst of errors over the entire frame. The loss of one or more slice groups can be easily concealed by neighboring macroblocks belonging to the slice groups which were received correctly. This facilitates a graceful recovery from errors. Moreover, since these errors are scattered over the entire frame, they are more easily concealed compared to the errors accumulated in a small region of the frame. However FMO does not lend itself to concealing loss of entire frames occurring due to signal loss intervals. Moreover, FMO also results in increased decoder complexity, and reduction in the coding efficiency of the encoder. Implementation and comparison of packet-level FEC strategies using Reed-Solomon (RS) codes at the link layer, and Raptor codes at the application layer for broadcast multimedia applications such as IPTV was presented in [58]. RS FEC coding and adaptive packetization of the video frame for recovery from erasures was proposed in [48]. Although FEC can be effective

in combating burst errors, it has an all or nothing performance. If the packet-loss rate exceeds the correction capability of the FEC code, almost nothing is available to the receiver. Enabling FEC to combat burst losses requires adapting the FEC redundancy to the length of the burst losses over the network [96], or interleaving packets (symbols) from multiple FEC blocks [70], which result in additional complexity and delay penalties. In Automatic Repeat Request (ARQ), the receiver can request retransmission of the lost packets. ARQ may be employed in conjunction with transport protocols such as the Transmission Control Protocol (TCP) for multimedia streaming over the Internet [103]. A relaxed form of ARQ, known as soft ARQ avoids retransmission of late data that would be discarded by the decoder [77], while hybrid ARQ which combines FEC and ARQ error control is proposed in [16]. However ARQ is not suitable for multimedia transmission scenarios where a return channel for requesting a retransmission is either not available, or is inconvenient to use, as in the case of broadcast and satellite communications. Moreover, retransmission of missing intra-frames in presence of a signal loss may result in unacceptable delays and bandwidth overheads. Several other approaches such as optimal inter/intra-mode switching for packet-loss resiliency [115], FEC [93], and FEC and pseudo-ARQ applied to layered coding [17] have been proposed to facilitate error-resilient video delivery over unreliable channels. In SVC and other layered video coding schemes, the source video is encoded as a base layer, and one or more enhancement layers to provide scalability in the spatial domain, the temporal domain, or in the signal quality. However, due to picture dependencies within a layered coded video sequence, any impairment on the base layer can result in impairments on all video layers of the sequence. As a result SVC by itself is not suitable for transmission over channels prone to burst errors, and has to be used in conjunction with other error-resiliency schemes in which the base layer can be unequally protected using FEC or ARQ. Multiple Description Coding is another popular approach for error-resilient video delivery [34, 104]. In MDC, several

independently decodable streams are generated spatially or temporally and transmitted over multiple paths [3]. If all the transmitted streams are received correctly, a high-quality video is reconstructed. However, in contrast to SVC, even if one or more encoded streams are lost, a low-quality video signal can still be reconstructed from the correctly received descriptions. The strength of the so-called multipath transport (MPT) approach lies in the assumption that spatio-temporally co-located segments of all the multiple descriptors are less likely to be impaired simultaneously when routed over disjoint channels [65]. However, it is not always possible to guarantee multiple dedicated channels between the source and client for the entire duration of video communication. Furthermore, given the non-deterministic arrival time of data over different channels, the receiver implementation of the MPT system tends to be highly complex. If all the multiple descriptions are to be transmitted over the same channel, these multiple bitstreams would need to be temporally interleaved in a manner that ensures that co-located segments belonging to different multiple descriptors are not impaired by the same burst or signal loss. Maximum separation between multiple descriptors can be achieved by sequentially sending one entire description followed by another on the same channel. However, this approach is obviously impractical because it introduces a transmission delay exceeding the duration of the entire video. Clearly, there is a need to design more feasible algorithms for dispersal of these multiple streams over a single channel.

In this chapter, we present a novel approach for error-resilient video delivery over a single bursty channel termed as Multiple Representation Coding (MRC). Similar to MDC, the MRC scheme relies on creating multiple independently decodable representations of the input, which are then streamed over a single channel. Instead of relying on path-diversity, a unique "GOP interleaving" approach is proposed to ensure that co-located frames of multiple representations are temporally dispersed,

thereby providing a graceful recovery from impairments caused by signal-loss intervals on the transmitted channel. The MRC scheme will be described and evaluated in the following sections.

3.2 Multiple Representation Coding

Figure 14 shows the block diagram of the proposed MRC system. The MRC scheme involves transmitting multiple downsampled representations of the source video to the receiver as a single stream. These multiple representations may be generated by spatial or temporal downsampling of the input sequence. The anti-aliasing filter prevents objectionable aliasing artifacts from appearing in the downsampled representations. The downsampled representations are then encoded using one or more video encoders. The encoded representations are then temporally dispersed on a single stream and transmitted over a single channel using the GOP interleaver.

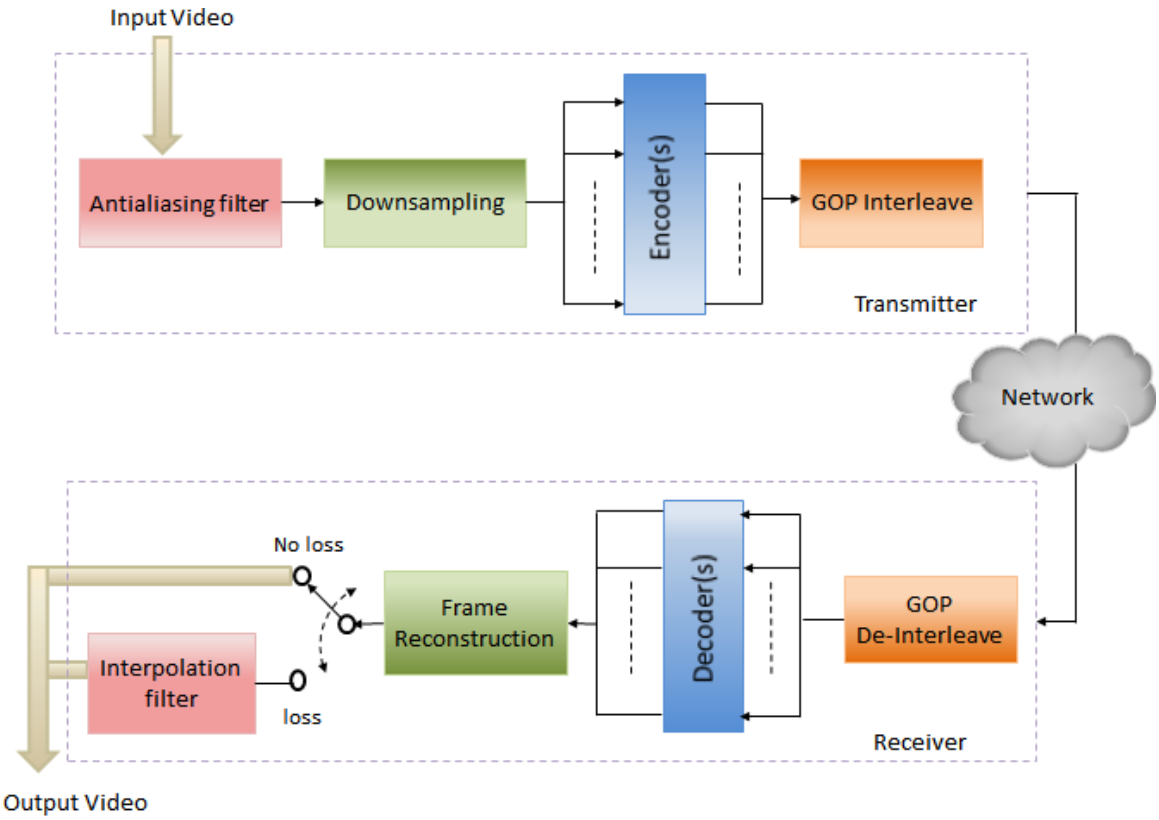


Figure 14: Block Diagram of the MRC System.

The receiver inverts all the operations performed by the transmitter. The received bitstream is deinterleaved to recover the multiple representations. These representations are then independently decoded using one or more video decoders. Decoded pictures from these representations are reassembled to recover the full-size frame. Missing pixels belonging to the impaired representations are approximated from the correctly-received pixels using a simple interpolation filter. As seen in Figure 14 the interpolation filter is not used if all the representations are received correctly. In the following few subsections, the various subsystems shown in the block diagram will be studied in detail.

3.2.1 Downsampling

The MRC scheme relies on creating multiple downsampled representations from the full-size source video. These multiple representations can be generated using several methods proposed in literature. Some approaches include polyphase spatial subsampling (PSS) [7], spatial subsampling by separating out top and bottom fields [60], temporal downsampling by separating out even and odd frames [3], oversampling and then downsampling transformed input frames [31], or using the FMO extension in H.264 as proposed in [88]. A " hybrid MDC " approach which segments the video in the spatial and the frequency domain to generate multiple descriptions is also proposed in [44]. After studying several methods to create multiple representations, we empirically determined to generate multiple representations by spatially or temporally downsampling each input frame, as seen in Figure 15. Further, spatial downsampling can be performed in one of the three configurations illustrated in Figure 15. Thus, it is possible to generate four multiple representations (4-MRC) as shown in Figure 15a or two multiple representations (2-MRC) as shown in Figure 15b and Figure 15c. The fidelity of error concealment of the MRC scheme against burst losses increases as the number of descriptions increases, whereas the error-robustness increases with

both: the number of representations, and the length of the representation segments interleaved in the video stream. However, this increased robustness comes at the cost of compression efficiency. As seen in Table 3, for the same value of the quantization parameter, the total bitrate of the transmitted stream increases with increasing number of representations. Thus, error-robustness and compression efficiency are clearly conflicting requirements which need to be carefully traded-off.

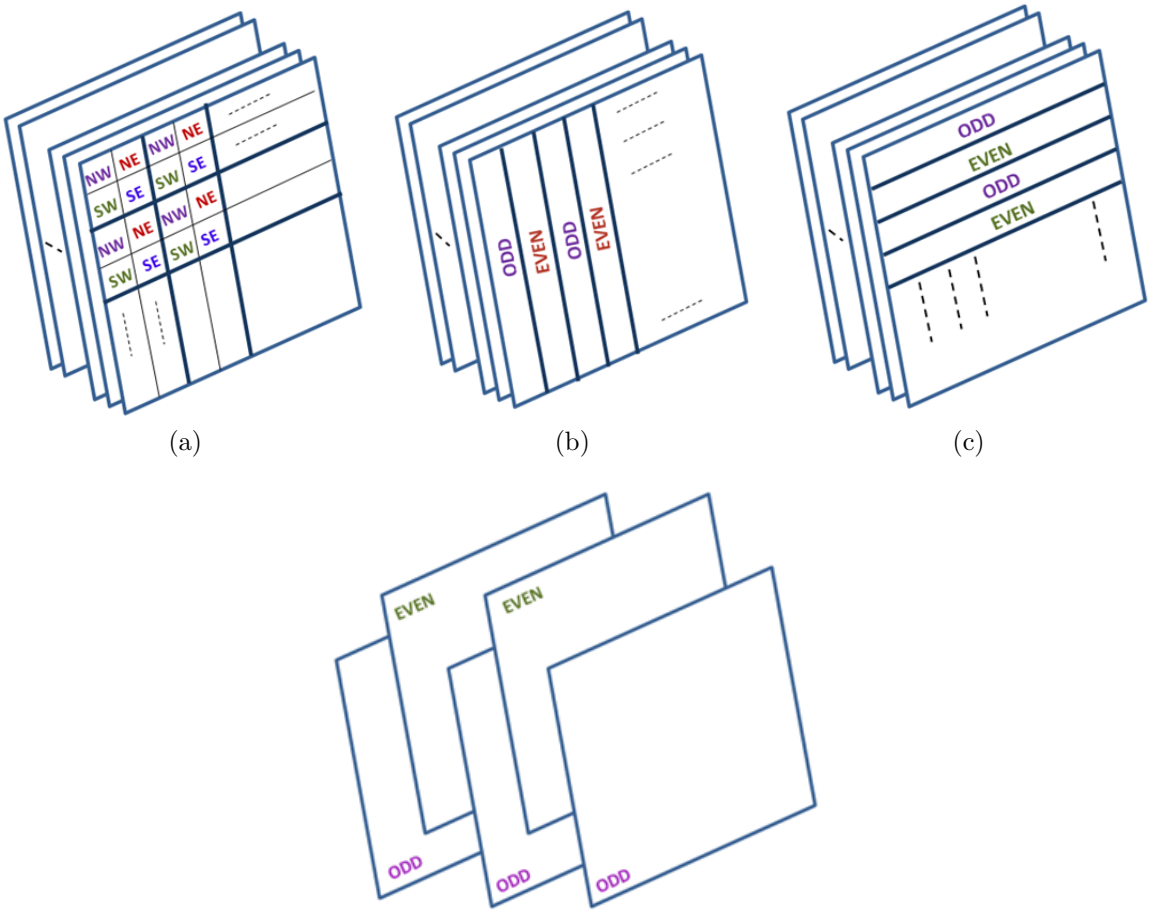


Figure 15: Spatial and temporal downsampling configurations for the MRC scheme.

3.2.2 Anti-Aliasing

The anti-aliasing filter prevents the appearance of aliasing artifacts in the downsampled representations. Figure 16a illustrates aliasing artifacts such as the jagged edges

Table 3: Total bitrate (kbps) for encoding the *Foreman* sequence [92] using different MRC configurations.

QP	Full-Size	2-MRC	4-MRC
22	921	1082	1352
27	442	541	670
32	221	274	338
37	112	139	171

of the building in the *Foreman* sequence. A three-lobed Lanczos downsampling filter [98] is used here. As seen in Figure 16b, the Lanczos-3 filter is very effective in preventing aliasing. Although the aliasing artifacts are visually unpleasant, the perceptual distortion introduced by aliasing artifacts is not accurately measured by fidelity measures such as the Peak Signal to Noise Ratio (PSNR). On the other hand, the picture modification occurring due to the low-pass filtering action of the anti-aliasing filter is readily reflected in the PSNR of each frame. As a result, the PSNR of a properly anti-aliased downsampled frame can be lower than that of an aliased downsampled frame as seen in Table 4. It can also be observed from Table 4 that in comparison to the aliased representations the anti-aliased representations can be compressed more efficiently. This again is a result of the low-pass filtering action of the anti-aliasing filter.

3.2.3 GOP interleaver

The key of the proposed MRC scheme is to temporally disperse data via the GOP interleaver. If the multiple representations are to be transmitted over a single channel, the GOP interleaver needs to ensure that corresponding sub-pictures of different representations are not impaired by the same burst loss. Typically, most sequences are encoded with a GOP length spanning approximately one second. Thus if the source frame-rate is 24 fps, each GOP contains 24 frames. If the full-size video is impaired by a signal loss, the decoder has to entirely rely on some form of error-concealment. But if the loss spans multiple GOPs, multiple seconds of the video sequence are lost,

(a) Anti-aliasing filter absent

(b) Anti-aliasing filter present

Figure 16: Spatial downsampling using the 2-MRC configuration (Figure 15c) (a) without Anti-aliasing filter and, (b) with Anti-aliasing filter.

and the decoder has to request retransmission. Retransmitting multiple frames of the sequence requires a reverse channel, and can also result in unacceptable delays. However in case of the MRC system, there are atleast two or more downsampled representations for every second of the video. If the GOP interleaver can ensure that the same burst loss does not impair all the spatio-temporally co-located segments belonging to multiple representations, then the error-concealment at the decoder can yield higher picture fidelity. The GOP interleaver separates the multiple representations of the 4-MRC configuration as shown in Figure 17. On careful examination of the transmission order shown in Figure 17, it can be seen that that temporally co-located GOPs of multiple representations are never adjacent to each other, and hence are less likely to be simultaneously impaired by the same burst loss.

Table 4: Effect of anti-aliasing filter on the PSNR and the bitrate of the *Foreman* sequence [92].

Anti-Aliasing Filter	PSNR(dB)	Bitrate(kbps)
Present	34.77	1352
Absent	39.55	1506

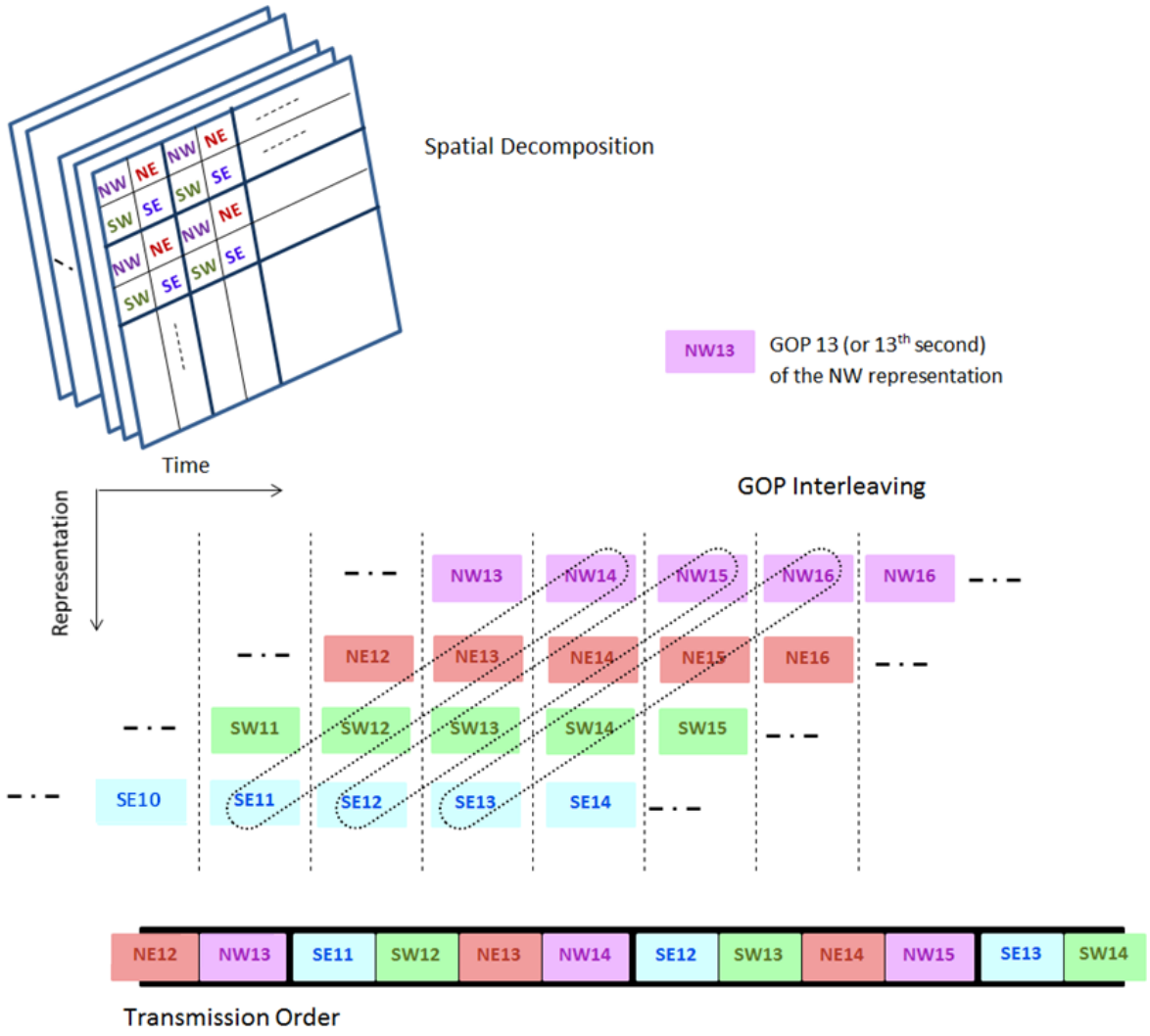


Figure 17: GOP interleaving for the 4-MRC configuration.

As seen in Figure 18, a burst loss which is four GOP-lengths long, would lead to a loss of four seconds, namely the 11^{th}, 12^{th}, 13^{th} and the 14^{th} second in the full-size scheme. On the other hand, a burst loss which is four GOP-lengths long would result in the loss of the SE11, SW12, NE13 and NW14 GOPs in the MRC scheme. This means that only one of the four representations for the same four second sequence (11 to 14) is completely lost. The missing pixels from the lost representation can be easily interpolated at the receiver. Figure 19 also shows the modifications in the GOP interleaver to accommodate a 2-MRC configuration. It is important to note that in contrast to a simple data or packet interleaving that may be performed on the lower

layers, the proposed GOP interleaver is highly video-centric. Since the lower layers are blind to the information carried on each packet, data or packet interleaving on the lower layers cannot guarantee that adjacent frames of the sequence are sufficiently dispersed on the transmitted stream, and hence cannot facilitate a graceful error recovery at the receiver. It is also important to note that the receiver has to wait for all the representations to arrive before attempting to reconstruct the full-size frames. This implies that the error-robustness of the MRC scheme comes at the expense of increased delay in reconstructing the received sequence.

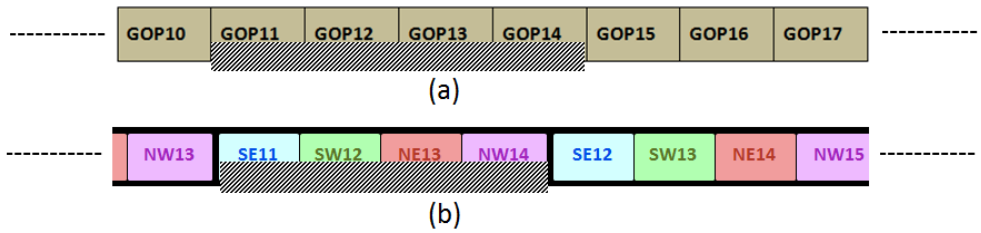


Figure 18: Simplified illustration of the effect of burst loss on (a) the full-size scheme, and (b) the MRC scheme.

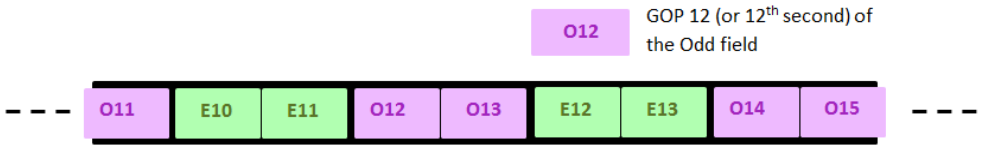
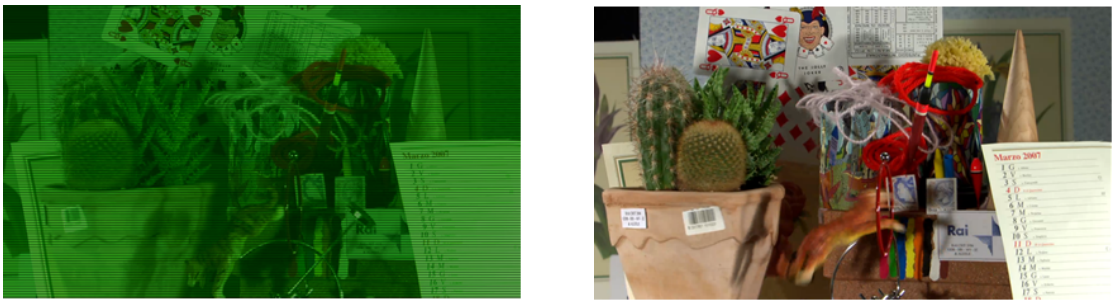


Figure 19: GOP Interleaving for the 2-MRC configuration.

3.2.4 Frame reconstruction and interpolation filter

The "frame reconstruction" block reorganizes the decoded representations to reconstruct the full-size frame. However, the reconstructed frame might be impaired due to loss of one or more representations. The missing pixels are approximated from the correctly received pixels using a very simple interpolation algorithm which uses the Lanczos-3 filter. Here, each correctly received representation is upsampled to the full-size frame using the Lanczos-3 interpolation filter. If multiple representations are received correctly then each representation is upsampled using the Lanczos-3

interpolation filter. These multiple upsampled representations are then averaged to approximate the pixels from the missing representation. In Figure 20a, one of the two representations encoded using the 2-MRC configuration (Figure 15c) is lost, resulting in a poor reconstruction at the receiver. However, by using the simple interpolation scheme described above, a good quality reconstruction can be obtained at the receiver, as seen in Figure 20b. The interpolation filter is not used if all the representations are received correctly.



(a) Interpolation filter absent

(b) Interpolation filter present

Figure 20: Reconstruction of a frame impaired by the loss of one representation in the 2-MRC configuration (a) without any error concealment and, (b) with error concealment.

3.3 Error resilient video delivery using the MRC scheme

For our experiments, we selected two common interchange format (CIF) sequences: *Foreman* and *Mobile*, and two 1080p sequences: *Sunflower* and *ParkJoy* recommended in the H.264 common conditions document [92]. All videos are encoded at four different bitrates using the x264 encoder [67]. The four different bitrates approximately correspond to QP values of 22, 27, 32 and 37 as recommended in [92]. For studying the performance of the proposed schemes we need to introduce bursts of packet losses. To introduce packet losses, the encoded bitstreams corresponding to both: the full-size video and the interleaved representations was "packetized" into fixed-size packets of 512 bytes. These packets were then dropped to simulate losses.

The length of the burst loss is important for estimating the distortion in the received video [62]. For comprehensive testing, we introduced single bursts of various lengths, approximately corresponding to a loss of zero to four GOPs of the sequence, as seen in Figure 21. For example, if a sequence is encoded with a GOP-size of 32, then BLL=$\frac{3}{2}$ would correspond to a loss of 48 frames. The simple *previous frame copy* method was used to recover lost frames of both the full-size and individual representations.

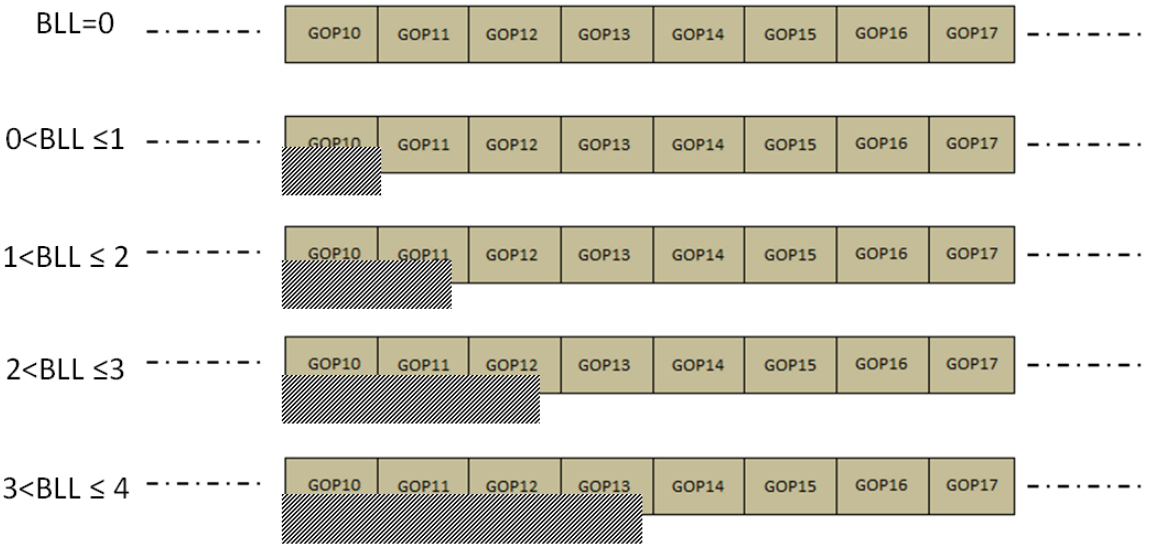


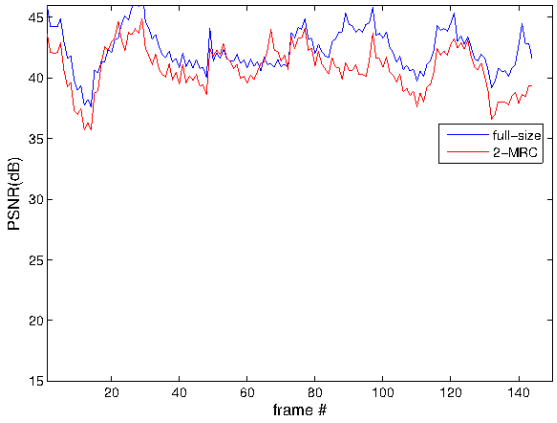
Figure 21: Illustration of burst losses of different lengths.

Figure 22 shows the per-frame PSNR for the *Sunflower* sequence encoded with average bitrate of around *2100 kbps* (for both full-size and 2-MRC). As seen from the Figure 22a the average PSNR for the MRC sequence is lower than the full-size sequence due to the redundancy of the MRC scheme. This loss in PSNR can be attributed to two factors. Firstly, there is a loss of PSNR due to the anti-aliasing filter as explained in Section 3.2.2. Secondly, the reduced coding efficiency of the MRC scheme implies that the individual representations have to be compressed more aggressively so as to stay within the target bitrate (the bitrate of the full-size scheme). Clearly, in absence of burst losses, the MRC scheme performs poorly as compared to the full-size scheme. However, as seen from Figures 22b and 22c, as the length

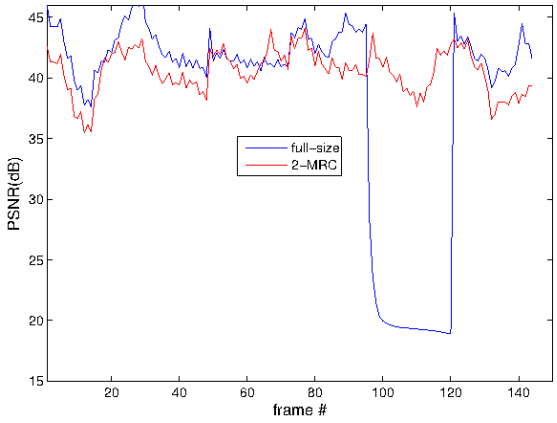
of the burst loss increases, the MRC scheme has a considerable advantage over the full-size scheme. The "valleys" in the PSNR-curves indicates loss of an entire frame. As can be seen from the plots, the "valleys" for the MRC scheme are either non-existent (Figure 22b), or occur for smaller durations (Figure 22c). It is well known that strong excursions in the picture quality are more visually annoying as compared to a picture with a slightly lower average quality but devoid of any abrupt changes. This implies that the MRC scheme can improve the overall viewing experience of the received sequence in the presence of burst losses or signal-loss intervals.

Besides the length of the burst loss, the location of the burst loss may also play a significant role in the nature of the introduced distortion. For the purpose of comprehensive testing, we generated multiple packet-loss traces. The location of the burst loss was randomly determined for each of these five traces. Figure 23 shows the average PSNR of the reconstructed sequence. As can be seen from Figure 23, the performance of the MRC scheme in comparison with the full-size scheme, improves as the length of the burst loss increases. We can also see that in comparison to the full-size scheme, the degradation in PSNR with increasing losses is much more graceful for the MRC scheme. Table 5 tabulates similar results for other sequences used in this experiment. The length of the burst loss in Table 5 is in the units of GOP-length. As seen from Table 5, the MRC scheme can facilitate a graceful recovery from long burst losses.

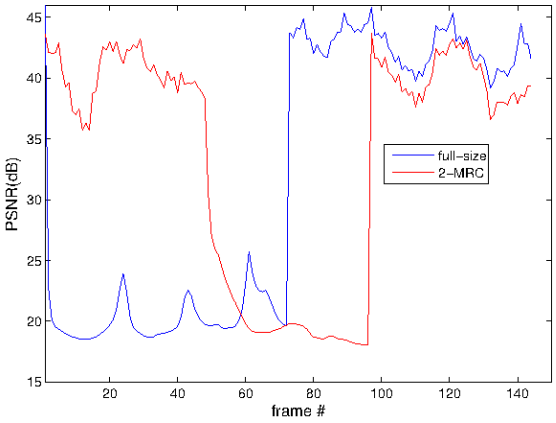
Figure 24, illustrates the effect of loss of multiple consecutive frames of the *Foreman* sequence on the visual quality of the received sequences. Figure 24a shows frame numbers 118, 124 and 127 of the Foreman sequence compressed at 442 kbps, and transmitted without any impairments. Since the previous-frame copy method is used to conceal the loss of frames, a "freeze-frame" effect is seen in Figure 24b. However, as seen in Figure 24c, a burst loss of the same length, and occuring at the same location does not result in a freeze-frame effect using the MRC scheme. This



(a) BLL=0 GOPs (0 packets), 42.33 dB, 40.78 dB



(b) 0<BLL≤1 GOPs (540 packets), 38.42 dB, 40.55 dB



(c) 2<BLL≤3 GOPs (2700 packets), 31.57 dB, 33.53 dB

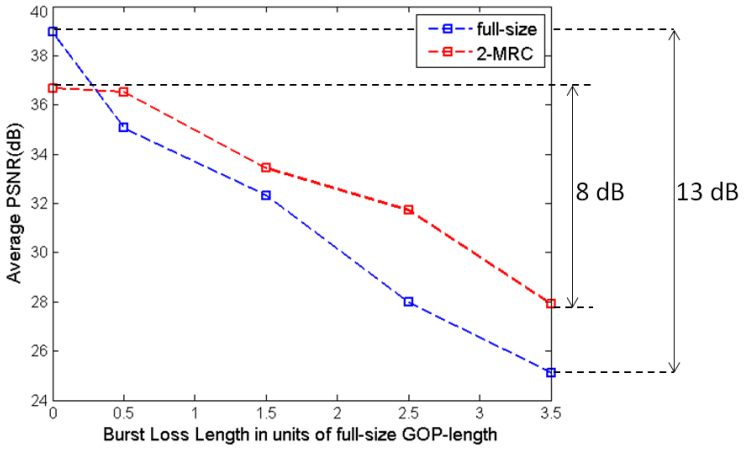
(d) The *Sunflower* sequence

Figure 22: The *Sunflower* sequence encoded with average bitrate of *6 Mbps* for both the full-size and the 2-MRC (Figure 15c) schemes. The caption of each sub-figure indicates the burst loss length (BLL) in units of GOP-length, and the average PSNR of the reconstructed sequences using the full-size scheme and the 2-MRC scheme.

is because only one of the interleaved representation is lost due to the burst loss. Pixels missing due to the lost representation are concealed by using the pixels of the correctly-received representation.

3.4 Adaptive GOP-length Multiple Representation Coding

As seen from *valleys* in the PSNR of the MRC curve in Figure 22c, the MRC scheme can also suffer from loss of entire frames for longer BLLs. Figure 25 illustrates the



(a)

(b) The *Sunflower* sequence

Figure 23: Average PSNR vs. the burst loss length in units of GOP-length for the *Sunflower* sequence. Average PSNR was obtained by averaging the PSNR over all frames of the sequence, and over all five loss traces, and over all four encoding bitrates.

Table 5: Average PSNR for the full-size (FS) and the 2-MRC schemes for different test sequences subject to various burst loss patterns.

Burst Loss Length	Foreman (220 kbps)		Mobile (270 kbps)		Sunflower (6 Mbps)		ParkJoy) (31 Mbps)	
	FS	**MRC**	**FS**	**MRC**	**FS**	**MRC**	**FS**	**MRC**
BLL = 0	32.93	30.11	25.17	22.15	42.33	40.79	32.86	29.57
0<BLL≤1	30.36	29.81	22.78	22.00	37.67	40.42	29.56	29.16
1<BLL≤2	27.63	27.42	20.58	21.01	35.62	37.20	26.20	26.78
2<BLL≤3	25.44	26.31	19.06	19.73	29.63	34.29	20.68	24.31
3<BLL≤4	21.92	23.04	16.59	17.54	25.92	31.41	19.14	22.05

limitation of the MRC scheme in the presence of long burst losses. As seen in Figure 25, if the burst loss impairs four or more GOPs in the stream, then frames from both representations corresponding to the 12^{th} and the 13^{th} GOP of the video can be impaired, and hence will degrade the error-concealment at the decoder. The error-robustness of the MRC scheme depends on the GOP-length which is used for encoding each representation. The length of the burst loss in Figures 26a and 26b is exactly the same. However, the GOP-length of each representation in Figure 26b is larger than the GOP-length of each representation in Figure 26a. As a result, in contrast to

(a) Full-size scheme without impairments. PSNR= 34.74 dB, **35.41 dB**, 34.79 dB

(b) Full-size scheme with impairments. PSNR= 34.74 dB, **24.48 dB**, 34.79 dB

(c) 2-MRC scheme with impairments. PSNR= 33.16 dB, **32.13 dB**, 34.55 dB

Figure 24: Illustration of the effect of loss of multiple consecutive frames of the *Foreman* sequence on the full-size scheme and the MRC scheme.

the Figure 26a, the frames belonging to the *even* representation and corresponding to the 12^{th} and the 13^{th} GOP of the full-size video are received unimpaired. Thus, the error-robustness of the MRC scheme can be improved by adapting the GOP-length of each representation to the measured burst loss over the network. Figure 27 shows the effect of varying the GOP-length on the error-robustness of the MRC scheme.

In Figure 27, the *Foreman* sequence was compressed using the 2-MRC configuration and varying GOP-lengths of 32, 38 and 48. All the sequences (full-size and 2-MRC) were compressed at a total bitrate of 183 kbps and were impaired by a single burst loss of length of around three seconds (or 96 frames). As can be seen from Figure 27, all the multiple representation coded sequences have considerably narrower *valleys* (or higher PSNR) as compared to the full-size encoded sequences in presence of the burst loss. Further, as the GOP length increases from 32 to 38, the width of the *valleys* reduces, and the *valleys* are completely absent for the sequence encoded with a GOP-length of 48. Thus as expected, increasing the length of each interleaved representation improves the error-robustness of the MRC scheme.

Figure 25: Limitations of the 2-MRC scheme in presence of burst lengths longer than 4 GOP-lengths.

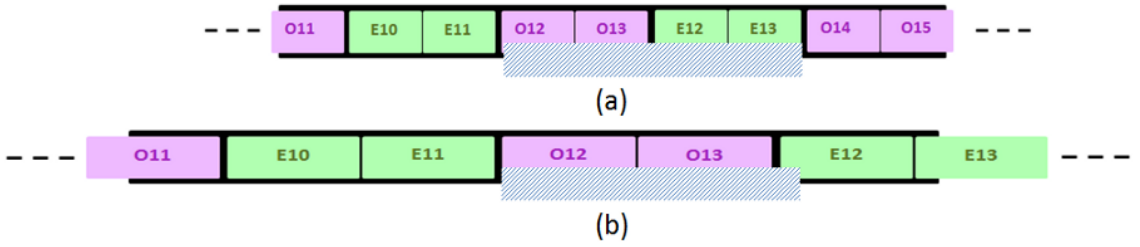
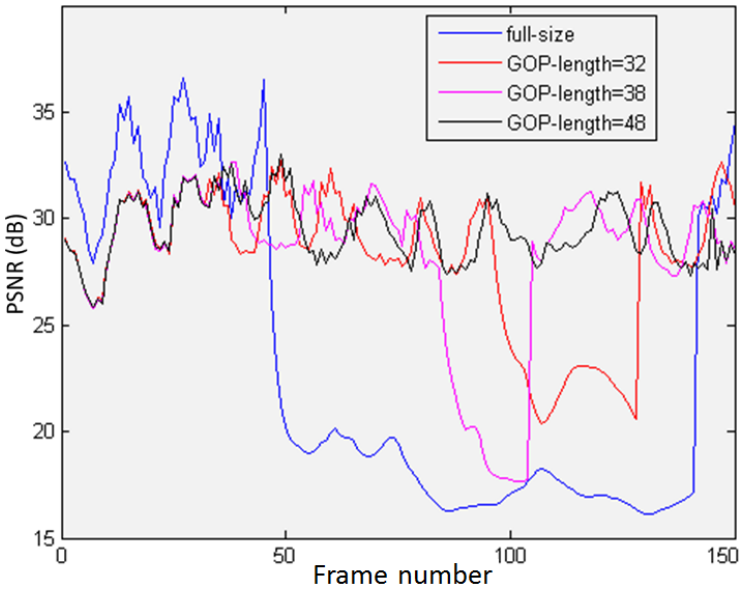
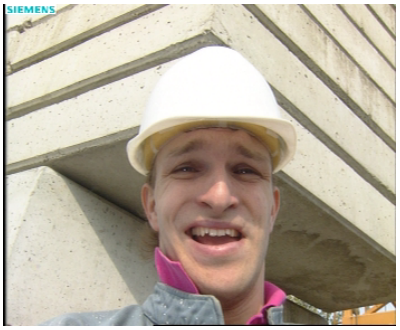


Figure 26: Effect of increasing the GOP-length on the error-robustness of the MRC scheme.

Figure 28 shows the block diagram of the adaptive GOP-length MRC system. The flow-graph of the GOP-length adaptation process is shown in Figure 29. The transmitter can receive some form of feedback from the receiver, or measure the network statistics in real time to determine the expected burst loss length. If the burst length is determined to be longer than a predetermined *threshold*, the GOP-length can be increased by a *factor*. The performance of the adaptive scheme, thus depends



(a) GOP-length variation



(b) The *Foreman* sequence

Figure 27: PSNR per frame for the *Foreman* sequence compressed using 2-MRC scheme at around 183 kbps.

on three design parameters: the *threshold*, which determines the amount of loss that can be tolerated by the system, the *factor* with which the GOP-length is increased, and the frequency with which the transmitter adapts the length of the representation segments. The *threshold* value determines the number of completely impaired frames that can be tolerated before the system begins the GOP-length adaptation. For example, for a *threshold* value of one, the system will increase the GOP-length as soon as a single frame-loss is detected. Taking into account the response vs. complexity trade-off, we fix the *threshold* value to four in our simulations. This implies that a loss of four consecutive frames triggers the GOP-length adaptation in our system. The *factor* with which the GOP-length is increased also determines the error-robustness of the system. As seen from the Figure 26, the overall error-robustness of the MRC scheme can be improved by increasing the GOP-length of each representation. However, increasing the length of the representation segments also increases the delay in reconstructing the received video sequence. Keeping the delay-robustness trade-off in mind, we place limits on our system to ensure that the GOP-length cannot be

increased by more than twice the full-size (or the original) GOP-length. Finally, the more frequently the transmitter responds to the receiver or network feedback, the faster it can adapt to the burst loss. In our system, we restrict the frequency of GOP-length adaption to a maximum of two times every GOP interval.

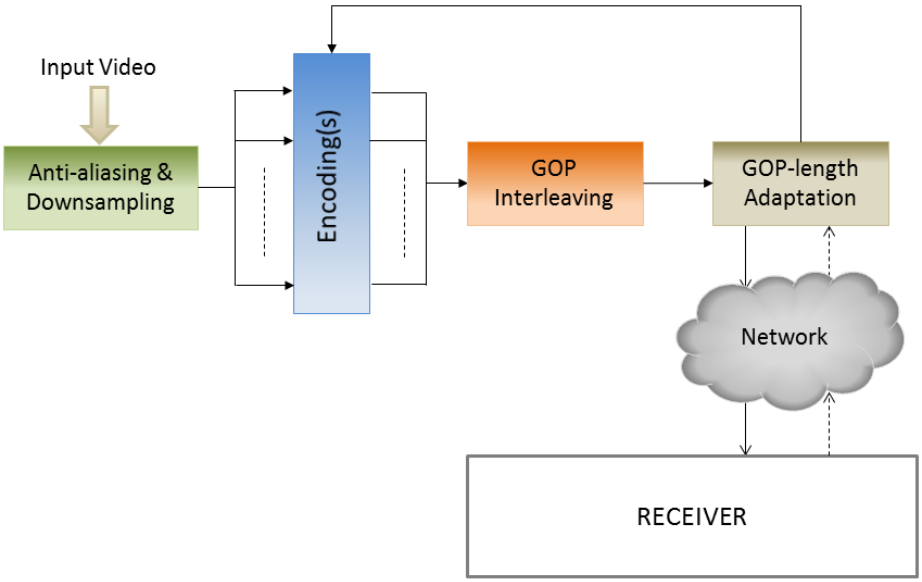


Figure 28: Block diagram of the adaptive GOP-length MRC scheme.

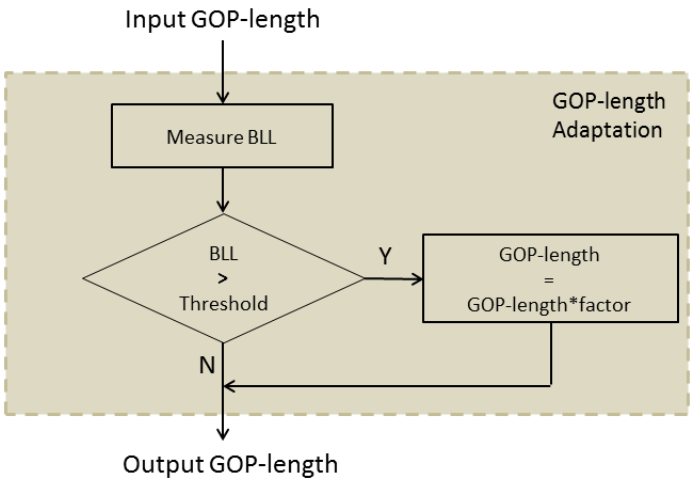


Figure 29: Flowgraph of the GOP-length adaptation process.

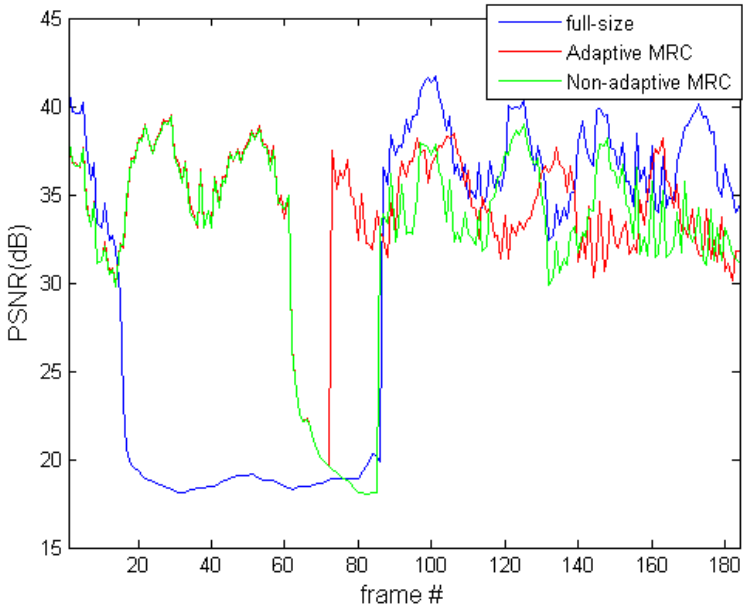
Figure 30 shows the per-frame PSNR for the *Sunflower* sequence, encoded with an average bitrate of around *1200 kbps*, for both: the full-size and the 2-MRC schemes. The BLLs indicated in Figure 30 are in the units of the original or the full-size

GOP-length. As seen from the Figure 30, the PSNR curves for the adaptive and non-adaptive MRC schemes overlap in the absence of burst losses, but differ significantly when a burst loss occurs. The adaptive MRC scheme responds to the loss of multiple frames by changing the GOP-length of the interleaved representations, and hence facilitates much better reconstruction fidelity as compared to the non-adaptive MRC scheme. As before, both: the adaptive and non-adaptive MRC schemes significantly outperform the full-size scheme.

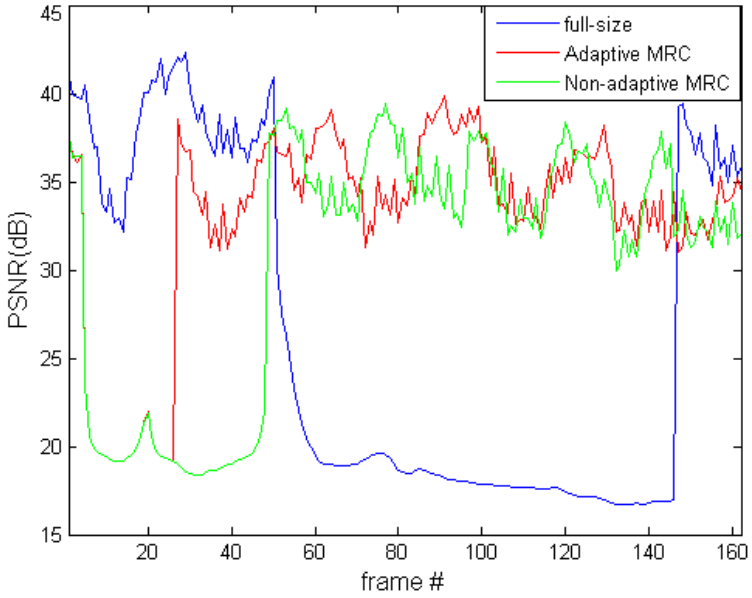
3.5 Comparison of MRC with the FEC coding

Forward error correction is a channel coding technique which adds redundancy to the transmitted data to enable error correction at the receiver. In a (n,k) FEC code, the original data containing k symbols, is transmitted as longer message containing n symbols. Thus an overhead of n-k symbols are added to facilitate recovery from packet losses at the receiver. Further, FEC codes may be systematic or non-systematic. In the systematic case, the symbols of the original message are included within the encoded message. To compare the performance of the MRC scheme with FEC, the full-size video is protected using the (n,k) systematic Raptor codes. The Raptor codes are a class of fountain codes [87]. They are also known as rateless codes, because the code length need not be fixed in advance. Use of Raptor FEC has been standardized by the 3rd Generation Partnership Project (3GPP) for use in multimedia broadcast and multicast services [12], by the Digital Video Broadcasting-Handheld (DVB-H) format [32], and has also been defined in the Internet Engineering Task Force (IETF) RFC 5053 [35]. Raptor FEC is employed here to improve the error resiliency of the full-size video.

Figure 31 illustrates the simulation setup for comparison of the fullsize stream protected with FEC, and the 2-MRC scheme. As seen in Figure 31a the amount of redundancy (n-k) added by the Raptor codes for GOP 12 and 13 of the full-size



(a) BLL=3 GOPs, 30.02, 32.82, 34.09

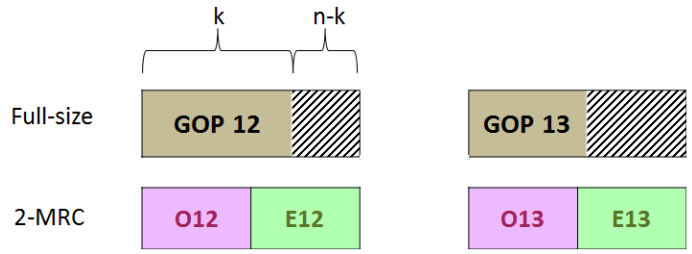


(b) BLL=4 GOPs, 26.40, 30.73, 33.01

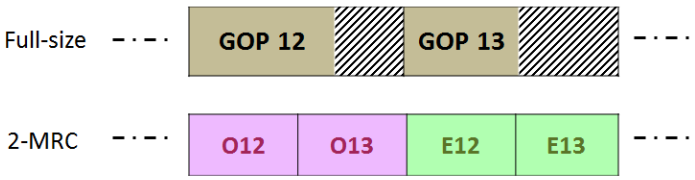
Figure 30: PSNR per frame for the *Sunflower* sequence encoded at a total bitrate of around *1200 kbps*. The caption of each sub-figure indicates the burst loss length in units of the full-size GOP-length, and the average PSNR for the full-size scheme, the non-adaptive MRC scheme, and the adaptive MRC scheme.

scheme is same as the coding overhead introduced by the 2-MRC scheme. Thus the total bitrate of the full-size scheme with FEC, and the MRC scheme can be equalized over a duration of one full-size GOP. Also, as seen in Figure 31b, due to the delay

introduced by the MRC scheme, GOP 12 of the full-size + FEC scheme can be decoded earlier than the MRC scheme. However the combined delay for decoding two GOPS: 12 and 13 for the 2-MRC scheme, and the full-size scheme protected by Raptor FEC is the same. Thus the delay for the full-size scheme with FEC and the 2-MRC scheme can be equalized over a duration of two full-size GOPs.



(a) Equalizing bitrates for GOP 12 and GOP 13 of the full-size + FEC and the 2-MRC schemes



(b) Equalizing delay for the 2-MRC and the full-size + FEC transmitted bitstreams

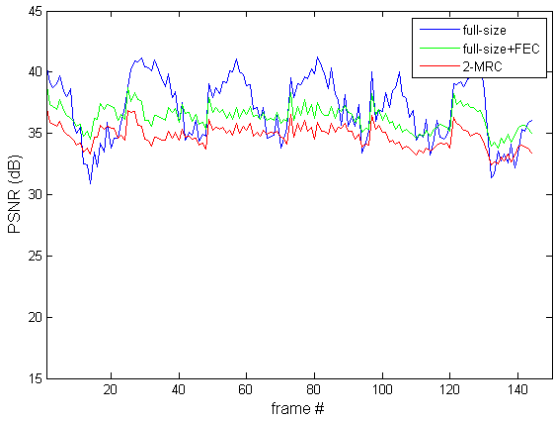
Figure 31: Simulation setup for comparison of full-size+FEC and 2-MRC schemes

Test sequences used for experiments in this section were same as those used in Section 3.3. All videos were encoded using the x264 encoder [67]. The Raptor FEC implementation described in [84] was employed here for protecting the full-size sequence. As described in Section 3.3, packet losses were introduced by "packetizing' the transmitted bitstreams into fixed-size packets of 512 bytes. These packets were then dropped to simulate losses. Each packet represents a symbol for the Raptor FEC encoder. Thus the symbol length for the Raptor FEC encoder was also set to 512 bytes. Other FEC encoder parameters such as number of symbols (k), and the overhead rate were decided on a per-GOP basis as described in Figure 31. For comprehensive testing, we introduce single bursts of various lengths, approximately

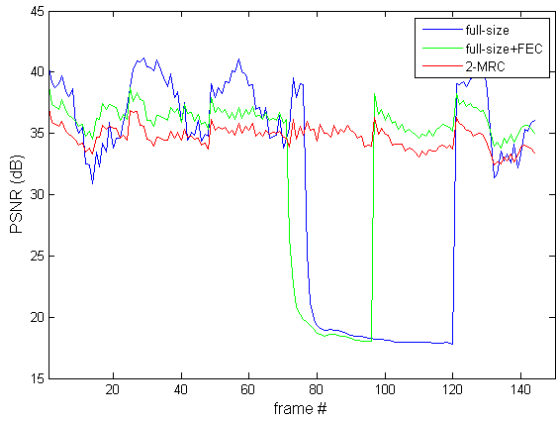
corresponding to a loss of zero to four GOPs of the sequence. The *previous frame copy* method was used to recover lost frames of both the full-size and individual representations. Figure 32 shows the per-frame PSNR for the full-size, 2-MRC and full-size+FEC schemes. All the transmitted bitstreams were encoded at approximately the same bit-rates. The performance of the all the three schemes is approximately the same in the absence of any burst losses. In the presence of burst losses of smaller length, the FEC scheme can improve the reconstruction of the full-size scheme as seen in Figure 32b. However, as evident from Figure 32c, as the burst loss length increases the error-correction capability of FEC degrades to a point where it does not offer any advantage over the full-size transmitted without any error-protection. As seen from Figure 32b and Figure 32c, the 2-MRC scheme can facilitate much better reconstruction fidelity as compared to the full-size scheme and the full-size+FEC scheme in presence of burst losses or signal loss intervals.

3.6 MRC for error-resilient delivery of HEVC sequences

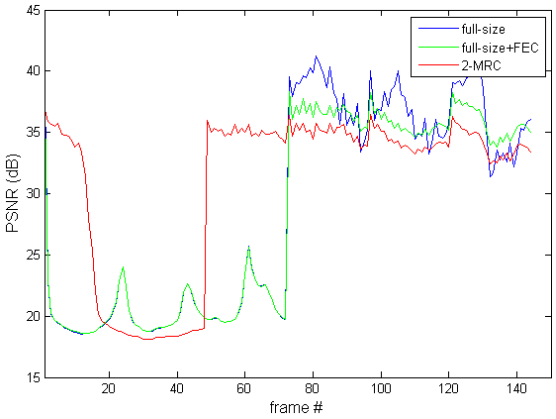
The MRC scheme relies on creating multiple independently decodable representations from the source video. In all the previous sections, the H.264/AVC video-coding standard, and the x264 encoder was used for encoding all representations. To study the utility of the MRC scheme in error-resilient delivery of H.265/High Efficiency Video Coding sequences, an HEVC encoder was used to encode the multiple representations in this section. H.265/HEVC is the new emerging international video coding standard, and is the successor to the H.264/AVC video-coding standard. The development of HEVC started when a joint call for proposals on video coding technology was issued by the Joint Collaborative Team on Video Coding (JCT-VC) January 2010 [41]. The JCT-VC was established by the ITU-T VCEG and the ISO/IEC MPEG with the aim of developing a new video coding standard to support beyond-HD frame sizes, and achieve greater coding efficiency as compared to AVC. The first edition of the HEVC



(a) BLL=0 GOP (0 packets), 37.03 dB, 36.29 dB, 34.74 dB



(b) 0<BLL≤1 GOPs (115 packets), 31.68 dB, 33.27 dB, 34.69 dB



(c) 2<BLL≤3 GOPs (574 packets), 28.70 dB, 28.28 dB, 30.85 dB

(d) The *ParkJoy* sequence

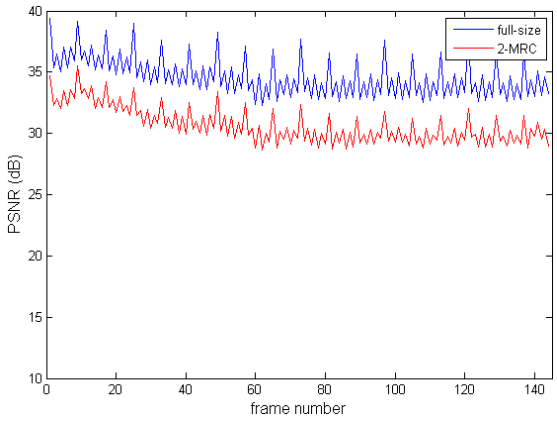
Figure 32: The *Sunflower* sequence encoded with average bitrate of *1200 kbps* for the full-size, the 2-MRC (Figure 19), and the full-size + FEC schemes. The caption of each sub-figure indicates the burst loss length in units of GOP-length and the average PSNR values for the full-size scheme, the full-size + FEC scheme, and the 2-MRC scheme.

standard was finalized in January 2013 [13], and has been reported to provide a bit-rate reduction of around 50% for perceptual quality equal to existing video-coding standards [89]. In this section, we will study the transmission of sequences encoded using this new standard over channels prone to burst losses and signal loss intervals.

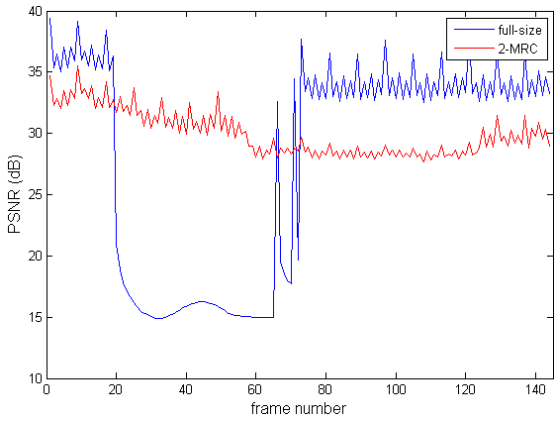
The database for the experiments performed in this section consists of four HD 1080p sequences: the *ParkScene* and the *Cactus* sequences recommended in the HEVC common conditions document [11], and the *Sunflower* and *ParkJoy* sequences

used in several previous experiments. The HEVC sequences were encoded using version 7 of the HEVC reference software (HM7) [47]. The transmitted bitstreams are impaired by burst losses in the same manner as described in Section 3.3. Figure 33, shows the performance of the 2-MRC scheme as compared to the full-size scheme. In Figure 33, transmitted bit-streams for both schemes were encoded at a total bitrate of around 31 Mbps. As seen in Figure 33a, the MRC scheme performs poorly in presence of no losses. As explained in Section 3.3, this loss in PSNR is due to the picture modification performed by the downsampling filter, and the coding overhead introduced by the MRC scheme. However, as the length of the burst loss increases, the 2-MRC scheme easily outperforms the full-size scheme. As seen from Figure 33, the "valleys" in the PSNR curve are either absent (Figure 33b), or are narrower (Figure 33c) than the full-size scheme in presence of burst losses.

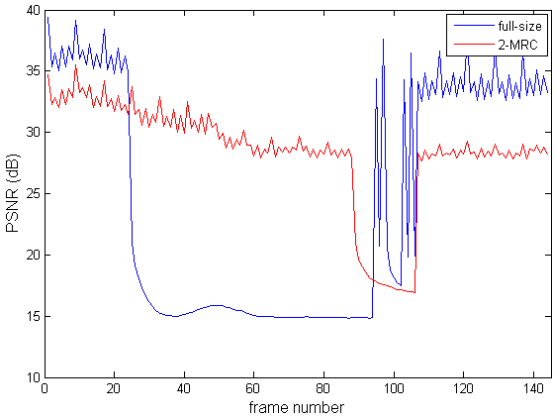
Since H.265/HEVC is an emerging standard, the HM reference software is currently the only freely available codec. Unlike x264, which supports an "average bitrate" mode, the HM encoder lacks the ability to generate HEVC bitstreams at a user specified bitrate. Instead, a user can only specify a value for the quantization parameter of the *I-frame* as an input to the HM encoder. However, it is impossible to predict the bitrate of the encoded bit-stream based on the QP of the *I-frame*. Moreover, the bitrate of the encoded full-size stream, and the total bitrate of the encoded downsampled representations at the same value of the QP are quite different from each other. This can be seen from Table 6, where the bitrates for the full-size and the 2-MRC schemes for all sequences except the *ParkJoy* sequence are different from each other. Due to this disparity in bitrates, direct comparison of the full-size and the 2-MRC schemes encoded with the same QP is not possible. To overcome this limitation, the sequences are encoded at four different values of the QP, namely 22, 27, 32 and 37. Cubic polynomial fitting of the log (base-10) bitrate against the average PSNR is then performed for each sequence. This helps to compute the Bjontegaard



(a) BLL=0 (0 packets) GOP, 34.59 dB, 30.64 dB



(b) 1<BLL≤2 (5800 packets) GOPs, 27.95 dB, 29.97 dB



(c) 2<BLL≤3 GOPs (9700 packets), 24.54 dB, 28.37 dB

(d) The *ParkJoy* sequence

Figure 33: The *ParkJoy* sequence encoded using HM, with an average bitrate of *31 Mbps*, or with a QP of *27* for the full-size, and the 2-MRC (Figure 19) schemes. The caption of each sub-figure indicates the burst loss length in units of GOP-length, and the average PSNR for the reconstructed sequences using the full-size, and the 2-MRC schemes.

Distortion (BD) metrics [8, 9]. BD-PSNR/Rate is a simplified method to compare the rate-distortion curves of reference sequence and the test sequence. In our case, the reference sequence is the full-size sequence and the test sequence is the 2-MRC sequence. The BD-PSNR measurement provides a single value in dB, computed from several test values at different bit rates. Figure 34 shows the fitted curves for the *Cactus* sequence for varying lengths of the burst loss. The curves seen in Figure 34 illustrate the average gain (or loss) in the PSNR provided by the 2-MRC scheme

when impaired by same burst loss as that of the full-size scheme. The BD-PSNR for different sequences impaired by burst losses of varying lengths can be seen in Table 7. A positive BD-PSNR indicates a gain in performance, since it corresponds to a gain in the PSNR at the same bit rate. As seen from Figure 34, and Table 7, as the length of the burst loss increases, the MRC scheme can facilitate a much higher reconstruction fidelity as compared to the full-size scheme.

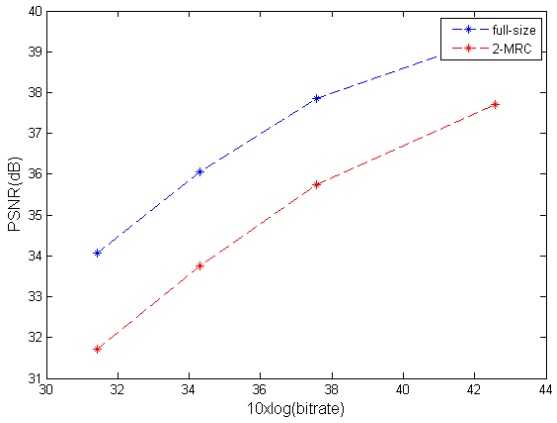
Table 6: Bitrate (Mbps) corresponding to the fullsize and 2-MRC schemes for various HEVC sequences compressed using different QP values.

QP	ParkScene		Cactus		Sunflower		ParkJoy	
	FS	**MRC**	**FS**	**MRC**	**FS**	**MRC**	**FS**	**MRC**
Qp=22	7.62	8.51	18.08	15.11	5.81	7.25	70.64	69.31
Qp=27	3.33	3.92	5.72	6.93	2.75	3.69	31.87	31.74
Qp=32	1.55	1.84	2.69	3.47	1.50	2.06	14.41	14.33
Qp=37	0.72	0.86	1.38	1.78	0.90	1.23	6.45	6.34

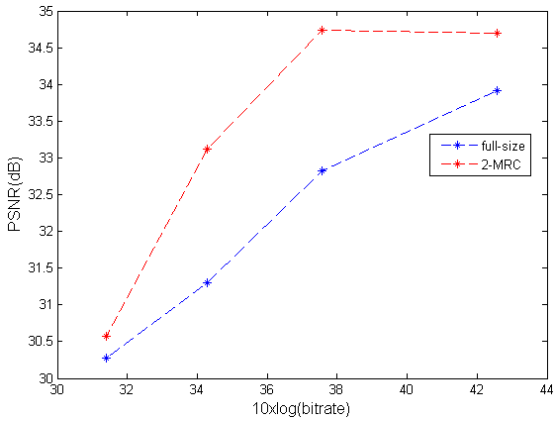
Table 7: BD-PSNR (dB) for various HEVC sequences impaired by losses of varying lengths.

BLL	ParkScene	Cactus	Sunflower	ParkJoy
BLL =0	-2.73	-2.12	-2.39	-3.44
0<BLL≤1	-0.48	-2.26	1.56	-1.28
1<BLL≤2	1.58	1.59	5.17	0.22
2<BLL≤3	3.16	2.24	6.00	1.45
3<BLL≤4	4.03	3.60	8.25	1.46

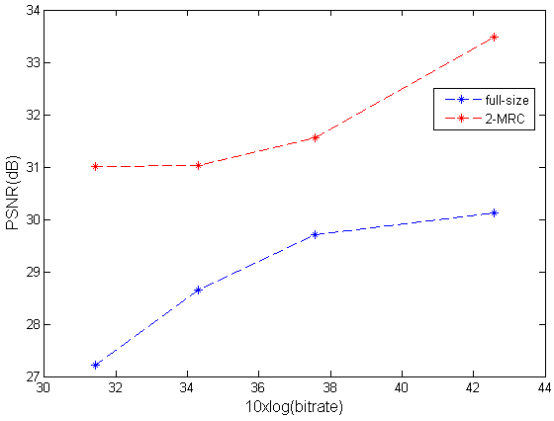
As seen from Section 3.2, the MRC scheme is encoder independent. Since the multiple representations are independently decodable, any codec can be employed to encode and decode the multiple representations. The HEVC standard was employed for encoding various sequences in this section, and an AVC encoder was employed in the previous sections. The MRC scheme was demonstrated to have a significant



(a) BLL=0 GOP, BD-PSNR=-2.10 dB



(b) 1<BLL≤2 GOPs, BD-PSNR=1.42 dB



(c) 2<BLL≤3 GOPs, BD-PSNR=2.40 dB

(d) The *Cactus* sequence

Figure 34: Calculation of BD-PSNR for the *Cactus* sequence with varying length of the burst loss.

advantage over the full-size scheme in the presence of burst losses, when employing either an HEVC or AVC encoder. However, we believe that the MRC scheme can potentially provide larger gains when the transmitted streams are encoded using the emerging H.265/HEVC standard. This hypothesis is based on observations made from Table 8, which shows the ratio of the *I-frame* to that of the non-reference B-frame (*b-frame*) for both: the HEVC and the AVC video coding standards. As seen from Table 8, the I:b ratio in HEVC sequences is over 3X the I:b ratio in AVC sequences. This implies, that *I-frames* constitute a larger percentage of the transmitted bitstream in case of HEVC as compared to AVC. As a result, for the same packet loss rate,

the *I-frames* of a HEVC sequence are more likely to be impaired by loss of packets as compared to the *I-frames* of the AVC sequence. Furthermore, due to coding dependencies, a loss of *I-frame* may result in the loss of the entire GOP of the encoded stream. Thus a loss of *I-frame* is much more catastrophic as compared to a loss of *P*, *B* or *b* frames. The MRC scheme can facilitate a graceful recovery from such impairments.

Table 8: Ratio of the *I-frame* to the *b-frame* in HEVC and AVC sequences

Sequence	QP	H.264/AVC I:b	H.265/HEVC I:b
ParkScene	22	545:1	1975:1
	27	1175:1	4023:1
Cactus	22	285:1	906:1
	27	500:1	1214:1
BQTerrace	22	1702:1	5849:1
	27	2560:1	7129:1

To verify this hypothesis, the *Sunflower* and *ParkJoy* sequences were encoded using HEVC and AVC encoders. The experiments previously described in this section were then repeated for these encoded sequences. The BD-PSNR measurements in Table 9 support our hypothesis. As seen from Table 9, the average gain in the PSNR of the reconstructed sequence, in the presence of burst losses of varying lengths is larger for the HEVC streams as compared to the AVC streams. This indicates that although the MRC scheme in itself is encoder independent, properties unique to certain encoded bitstreams such as H.265/HEVC can benefit more from the error-recovery facilitated by the MRC scheme.

3.7 Summary

In this chapter, we introduced a novel approach for error-resilient video delivery known as Multiple Representation Coding. The MRC scheme employed a novel GOP

Table 9: Comparison of MRC scheme used in conjunction with the HEVC and AVC video coding standards

BLL	*Sunflower*		*ParkScene*	
	H.265/HEVC	**H.264/AVC**	**H.265/HEVC**	**H.264/AVC**
BLL =0	-2.39	-2.51	-2.73	-2.45
0<BLL≤1	1.56	1.56	-0.48	0.15
1<BLL≤2	5.17	3.49	1.58	0.93
2<BLL≤3	6.00	4.36	3.16	2.98
3<BLL≤4	8.25	5.53	4.03	2.90

interleaver, to temporally disperse the multiple representations of the source video over a single stream. We demonstrated that in contrast to full-size encoding and transmission of the entire video sequence, it is possible to achieve a graceful recovery from burst errors and signal-loss intervals by using the MRC scheme. Further, we exploited the inherent flexibility of the MRC scheme, by adapting the GOP-length of the individual representations in response to the length of the burst loss over the network. Experimental results demonstrated that the adaptive MRC scheme can deliver the video with higher fidelity as compared to the non-adaptive MRC scheme and the full-size scheme in the presence of burst or signal losses. We also performed a study to compare the error-recovery facilitated by the MRC scheme and the error-correction facilitated by the Raptor FEC coding. Experimental results demonstrated that the MRC scheme can result in much better reconstruction fidelity as compared to FEC in presence of burst losses, and signal loss intervals. Finally, we investigated the effectiveness of the MRC scheme in facilitating error-resilient delivery of sequences encoded using the emerging H.265/HEVC video-coding standard. It was determined that the MRC scheme can greatly improve the reconstruction fidelity of HEVC sequences impaired by burst losses and signal loss intervals. The findings in this chapter were published in [54], [52] & [55].

CHAPTER IV

ERROR-RESILIENT DELIVERY OF REGION OF INTEREST VIDEO USING MULTIPLE REPRESENTATION CODING

4.1 Introduction

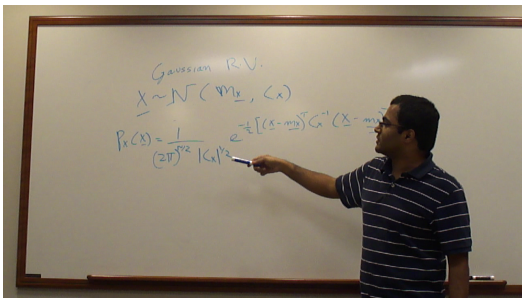
Distance education or distance learning is set to take place when the teacher and students are separated by physical distance, time, or both [74], and education is delivered remotely using technology (i.e., voice, video, data, and print) [111]. Distance education in some form or the other has been practiced for many decades. However, the recent Massive Open Online Courses (MOOC) initiatives such as Coursera [22] and edX [28] launched by major universities, promise an effective alternative to conventional classroom learning for students. MOOC platforms also vastly increase the ability of the universities offering such courses in reaching a larger audience [75]. Distance learning can be facilitated either synchronously or asynchronously [108]. In synchronous learning, all the participants are present at the same time, and hence tools such as live chat or video conferencing can be employed to overcome the geographical separation. If the information provider and the learners are separated by time, then they can engage in asynchronous learning facilitated via emails, message board forums, and pre-taped video lectures. However, as discussed in Chapter 2 and Chapter 3, enabling high-quality video communication to facilitate a usable distance learning experience to the remotely located learners, can be impaired due to bandwidth constraints and packet losses over the underlying network. Video connection problems have been cited as one of the barriers to enabling a seamless distance learning experience [99]. This problem is further aggravated due to the broad reach of

the MOOC platforms. One article has reported that students from "Paraguay to Pakistan, India to Ghana, Indonesia to Iraq, to Morocco, to Nigeria, to Australia, to Serbia" participated in a distance learning course offered on Coursera [106]. However, availability of reliable broadband connections in rural regions within the US, and within some of the developing nations mentioned above cannot be guaranteed. In [95], MPEG-1 video content was transmitted over satellite links, and transmission speed of 2.5 Mbps was determined to be the lower threshold for enabling distance learning and tele-education. The effect of competing load traffic on multimedia distance learning over IP networks was studied in [30]. The authors reported that audio and video content added to text increases the learning effectiveness for learners when the communication quality is high. However they found that degradation in the audio and video quality due to increasing traffic over the transmission network decreases the learning effectiveness of the system to a point where text-only transmission is preferred.

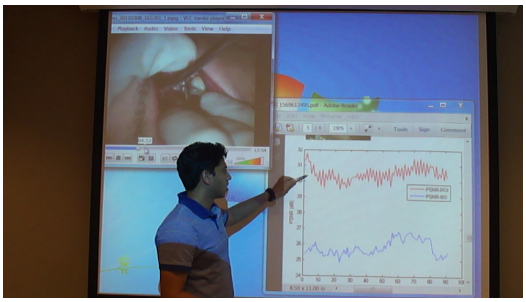
Clearly, delivering a functional distance learning experience to the remote learners requires application-specific optimizations such as those defined in the ASVCD toolkit. To that effect, we propose employing ROI-based video coding for encoding the critical regions of the frame at a higher quality as compared to the background. Further, we propose unequal protection of the ROI for ensuring reliable delivery of the high-quality ROI.

Figure 35 presents an experimental database of sequences representing a typical asynchronous distance learning scenario. All sequences in this database were 720p, 30 fps sequences containing clips involving a speaker set against a stationary or a moving background, a speaker writing on the white board, and a speaker presenting content projected on a screen.

Videos typical to a classroom learning scenario (Figure 35) generally contain a very well-defined ROI. This ROI could be the handwritten content on the white board, the



(a) Equation



(b) Presentation

(c) Speak

(d) Interview

Figure 35: Database of distance learning videos.

content projected on the screen or the gestures and face of the speaker. As seen in Chapter 2, ROI video coding is one possible solution to facilitate transmission of such sequences over bandwidth-constrained resources. By encoding the ROI with more number of bits as compared to the rest of the frame, the ROI can be delivered with a higher visual quality to the end-user. Several techniques to facilitate ROI-based video coding were reviewed in Chapter 2. An iterative approach was used to assign the highest possible compression level to the BG, and the lowest possible compression level to the ROI without exceeding the target bitrate in [15]. Several other approaches employ the FMO tool defined in the H.264/AVC standard for enabling ROI support [26, 100]. Besides telemedicine [53, 81], ROI-based video coding has proven useful in a broad range of applications such as surveillance [5], and video telephony [63]. A few instances of employing the notion of ROI in distance-learning videos can be seen in the literature [6, 45]. However, these approaches focus on methods to extract and track the ROI, rather than addressing ROI-based encoding of the videos. In spite of the lack

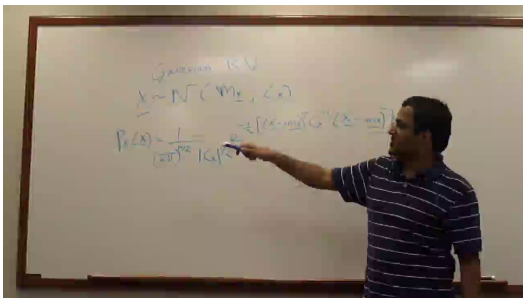
of sufficient evidence in favor of ROI video coding for provisioning distance learning, the ability of ROI-based video coding solutions in delivering the critical regions of the frame with a higher quality was comprehensively demonstrated in Chapter 2. Moreover, a recent study has also shown that low-bitrate videos encoded with a ROI support have a higher perceptual quality compared to uniformly encoded videos [43]. Clearly, ROI-based video coding in low bitrate scenarios can increase the bandwidth perceived by the video delivery application, and can prove very useful in enabling distance learning.

As mentioned previously, besides bandwidth limitations, the quality of experience delivered to the end-user rapidly degrades due to presence of burst losses caused by congestion, and signal losses over wireless and cellular networks. As seen in Chapter 3, several techniques such as FEC [48], FMO in H.264/AVC [59] and MDC [34, 104] have been proposed to facilitate error-resilient video delivery over unreliable channels. However, all these approaches face limitations in the presence of burst errors or signal-loss intervals. The MRC scheme was shown to address these limitations in Chapter 3. Instead of relying on path-diversity, the MRC approach temporally disperses multiple representations on a single transmitted stream to facilitate a graceful recovery from impairments caused by burst and signal losses. However, rather than attempting error-resilient delivery of the entire coded video sequence, it might be more desirable to ensure error-free delivery of just the ROI in case of applications such as distance learning. In [40], the picture is divided into foreground and background sub-pictures. The macroblocks in the background sub-picture are more strongly quantized as compared to the MBs in the foreground sub-picture. Further, unequal error protection (UEP) and slice interleaving is employed at the packet level to improve the error resiliency and concealment in the foreground sub-picture. Although a typical ROI is not established in [94], the frame is partitioned into three slices of "low", "medium", and "high" importance using the FMO tool in H.264, and then unequally protected

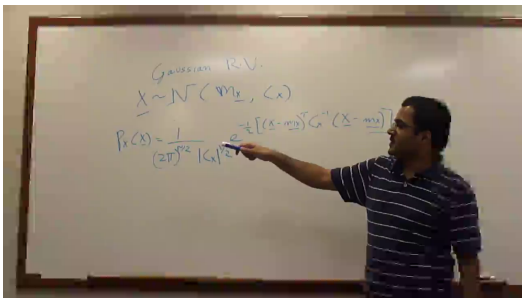
using Reed Solomon codes. Instead of using error correction codes for providing UEP to the ROI, Multiple Description Coding is employed in [72]. Here, multiple regions of interest are established using the set partitioning in hierarchical trees algorithm, and each ROI is placed on a separate independently decodable description. If all the descriptions are received error-free, all the regions of interest are recovered. However, the ROI encoded in any lost description cannot be recovered using this scheme. In [4], the FMO tool in H.264 is used to split the video frame in two foreground slices and one background slice. The foreground slices encode the ROI in a checkerboard pattern. Further, the foreground slices are protected with a stronger error-correction code as compared to the background slice. UEP for ROI-encoded JPEG 2000 images was also proposed in [113]. Given the limitations of existing error-resiliency techniques described in Chapter 3, we propose employing the MRC scheme to facilitate error-resilient delivery of the ROI in presence of burst losses and signal loss intervals.

4.2 Region of interest video coding

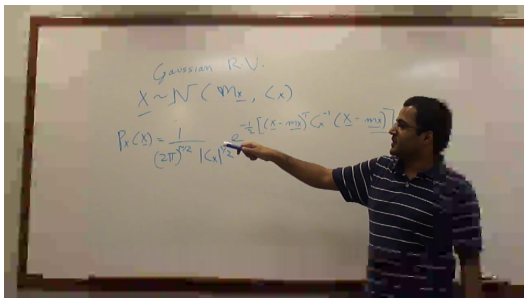
A flexible, interactive, and low-complexity method to introduce ROI support in coded videos was already introduced in Chapter 2. As seen in Chapter 2, the ROI is established by modifying the quantization parameter of each MB in the frame. Since QP is inversely proportional to the bitrate, more bits can be assigned to the ROI, by decreasing the QP of all the MBs occupied by the ROI. Similarly, fewer bits can be assigned to the BG by increasing the QP of all the MBs in the BG. To establish the ROI, a user-defined, positive or negative 'QP-offset' is assigned to each MB of the frame. This QP-offset is added on top of the QP decision made by the rate-control algorithm of the encoder. Figure 36 shows the effect of using different QP-offsets. Further flexibility is introduced in the system by allowing the ROI to occupy five different locations in the frame: northwest, northeast, southwest, southeast and center. Figure 37 shows frames of videos encoded with the various possible ROI-locations



(a) QP-offset=0 (No ROI)



(b) QP-offset=5 (Moderate ROI)



(c) QP-offset=10 (Strong ROI)

Figure 36: Effect of changing the QP-offset value on the central-ROI, and the BG quality of the *Equation* sequence.

in our implementation. The quality and the location of the ROI is signaled by the remote user.

4.3 Multiple Representation Coding

Multiple Representation Coding was added to the ASVCD toolkit in Chapter 3. MRC was shown to be effective in facilitating error-resilient video streaming over channels prone to burst losses and signal-loss intervals in Chapter 3. To briefly review Chapter 3, the MRC scheme involves creating multiple downsampled representations from the source video as seen in Figure 15. These multiple representations are then encoded and interleaved on a single transmitted stream using the GOP interleaver. As seen in Figure 17, the GOP interleaver temporally disperses the multiple representations in a manner that ensures that temporally co-located GOPs of different representations are never adjacent to each other, and hence, are less likely to be impaired by the same burst loss. Figure 38 demonstrates the effectiveness of the MRC scheme over

(a) SE-ROI

(b) NW-ROI

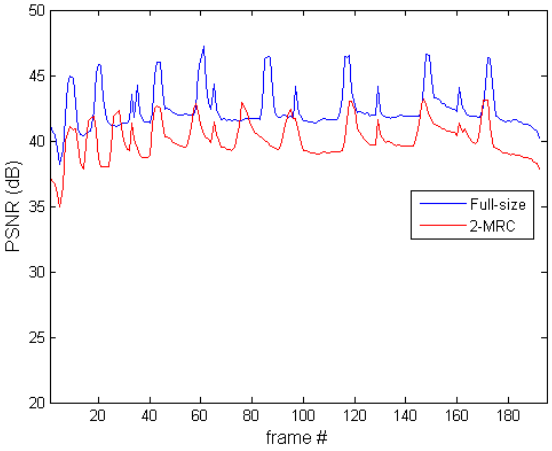
(c) Center-ROI

Figure 37: Some possible locations of the ROI in our system.

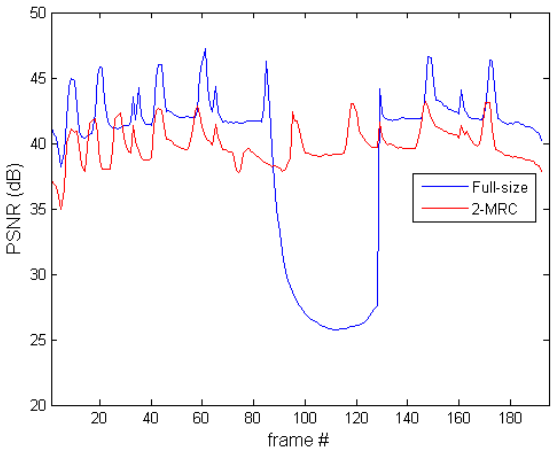
full size encoding and transmission of the video in presence of burst losses or signal loss intervals. As seen in Figure 38, as the length of the burst loss increases, the valleys in the PSNR curves of the 2-MRC scheme are either absent (Figure 38b), or are narrower than the full-size scheme (Figure 38c).

4.4 Unequal protection of the ROI using the MRC scheme

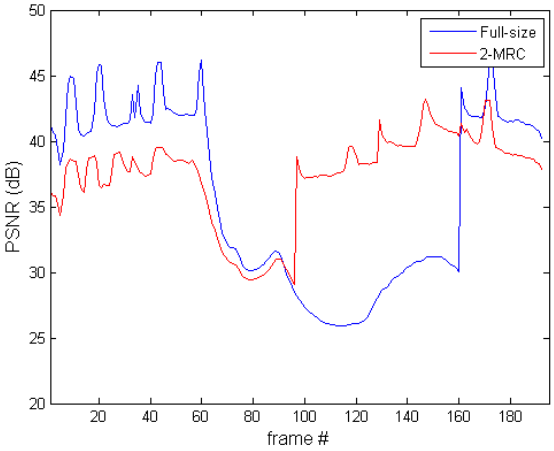
As seen in Chapter 2, although the ROI is the critical portion of the frame, the BG is often used by the viewer to gain overview information about the scene. However, in the event of burst or signal losses, it is more important to ensure error-resilient delivery of the ROI of the frame, over the BG. Video transmission over lossy channels can be improved by prioritizing error-free delivery of the ROI over rest of the frame. Here, we propose using the MRC scheme for achieving that goal. As described previously, in the MRC scheme, the source video is downsampled to generate two independently



(a) BLL=0 GOP (0 packets), 42.40 dB, 40.12 dB



(b) 0<BLL≤1 GOPs (132 packets), 39.18 dB, 39.88 dB



(c) 2<BLL≤3 GOPs (662 packets), 35.74 dB, 37.35 dB

(d) The *Interview* sequence

Figure 38: The *Interview* sequence encoded with average bitrate of *1100 kbps* for both the full-size and the 2-MRC schemes. The caption of each sub-figure indicates the burst loss length in units of GOP-length and the average PSNR for the full-size, and the 2-MRC schemes.

decodable representations. For a ROI-encoded video using the 2-MRC configuration (ROI+2-MRC), two independently decodable, downsampled representations are generated as shown in Figure 39. These representations are then encoded using the ROI encoder described in Section 4.2 and then interleaved via the GOP-interleaver in the order shown in Figure 40.

As seen from Figure 39 and Figure 40, in contrast to the BG, the ROI gets encoded

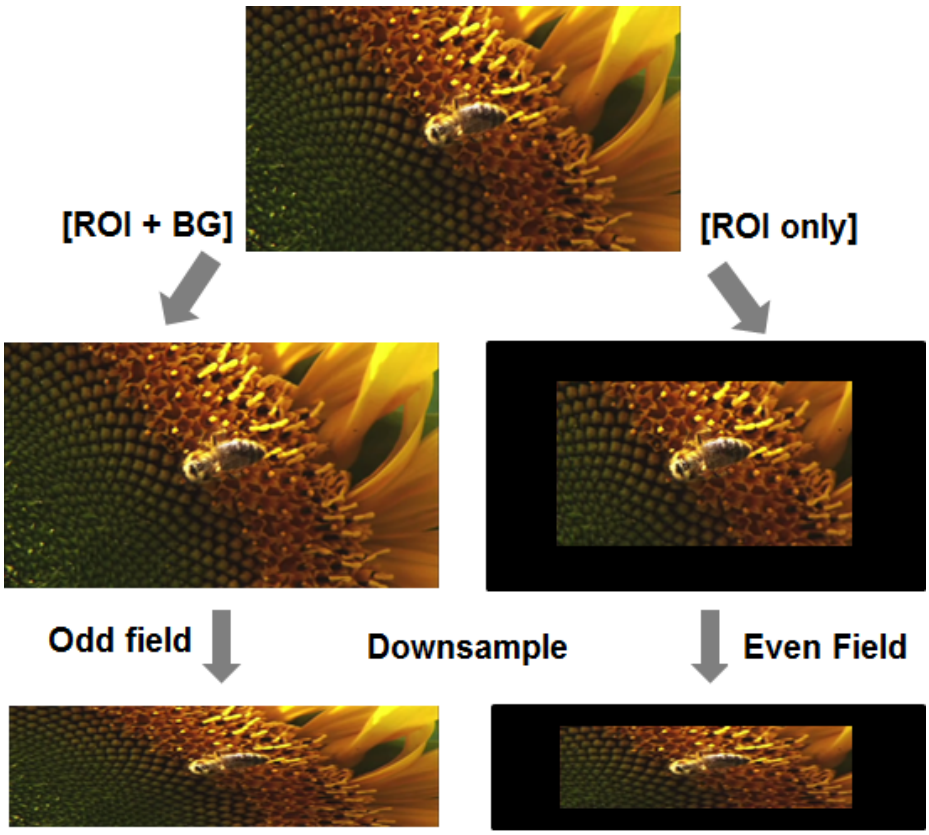


Figure 39: Process of generating two downsampled representations from the source video for the ROI+2-MRC scheme.

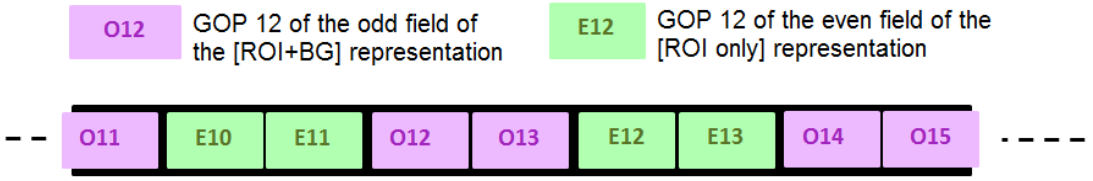


Figure 40: GOP Interleaving for unequal protection of ROI using the 2-MRC configuration (Figure 15c).

twice (once by each representation), and hence is more likely to be recovered with a higher fidelity as compared to the BG. The receiver can expect the following four cases.

Both representations are received: In this case, both the received representations are used to reconstruct the ROI. It is important to note that since the BG is downsampled and encoded by only one representation, the entire BG of

the source frame is *never* available to the receiver even if all representations are received correctly. Thus, the BG always needs to be upsampled to the full size at the receiver.

Only one representation - *ROI+BG* is received: In this case, the downsampled ROI and the BG have to be upsampled at the receiver to reconstruct the full-size frame.

Only one representation - *ROI only* is received: In this case, only the ROI can be reconstructed using interpolation. Since no BG information is available for reconstruction, the lost BG is concealed using the background information from the frames received previously. This case clearly demonstrates that as compared to the BG, the ROI is unequally protected using MRC. One should note that transmission of the entire ROI, and just the downsampled BG, further enhances the quality of the ROI at the expense of the BG.

Both the representations are lost: Finally, in this case, no information about the frame (ROI or BG) is available to the receiver, and hence the receiver has to rely on some form of error-concealment such as "previous frame copy" or "motion vector copy" to generate the display frame.

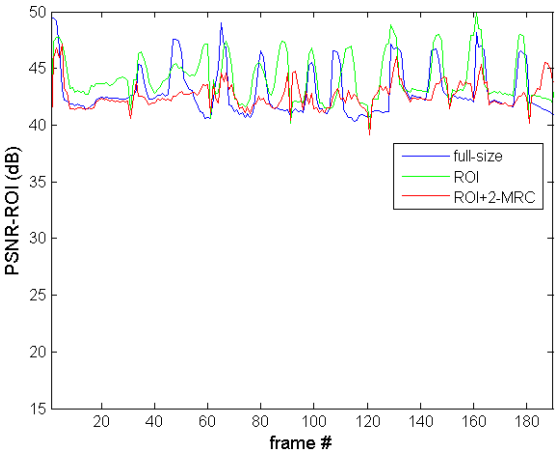
4.5 Results and Discussions

Videos from the database shown in Figure 35 were used for the purpose of simulations in this study. Besides distance learning, more generic applications such as mobile and wireless streaming can also benefit from the ROI+MRC scheme. Another set of videos consisting of the *Carphone*, *Miss America*, *Sunflower* and the *ParkJoy* sequences [92] were also employed here, for simulation purposes. For studying the performance of the proposed scheme, the videos were encoded using the ROI video-coding algorithm described in Section 4.2, and then subject to bursts of packet losses.

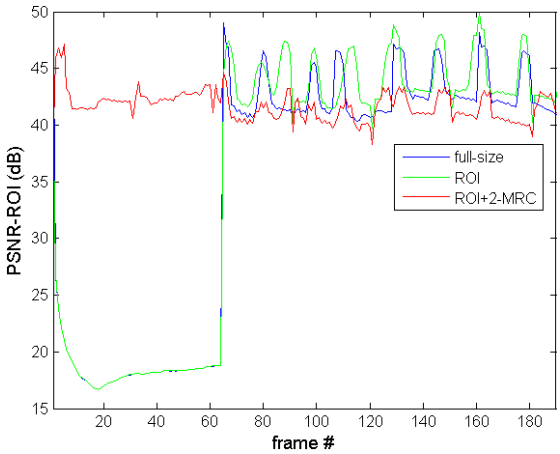
The simulation setup for introducing losses in the transmitted stream is the same as the one described in Section 3.3. Thus, to introduce packet losses, the ROI encoded bitstreams corresponding to the full-size video, the full-size video with ROI support, and the ROI-encoded video protected using MRC, were packed into fixed-size packets of 512 bytes. These packets were then dropped to simulate losses. For comprehensive testing, single bursts of various lengths, approximately corresponding to a loss of zero to four GOPs of the full-size sequence were introduced. A simple "previous frame copy" method was used to conceal lost frames of both the full-size and individual representations.

Figure 41 shows the per-frame PSNR in the ROI (PSNR-ROI) for the *Presentation* sequence encoded with average bitrate of around *2700 kbps* for all the three schemes: full-size encoding (uniformly encoded without ROI), full-size encoding with ROI, and ROI-encoding with 2-MRC. In the absence of burst losses (Figure 41a) the PSNR-ROI of the full-size video with a ROI sequence is higher than the other two schemes. This is expected because of the absence of ROI in the uniformly encoded full-size sequence, and the losses introduced by the MRC scheme in the ROI+2-MRC method. However, as seen from Figure 41b and Figure 41c, as the loss length increases the valleys in PSNR-ROI curve for the ROI+2-MRC scheme are significantly narrower than the other two schemes; thus improving the reconstruction fidelity of the received ROI. Further as seen in Figure 42, the ROI+2-MRC scheme results in much graceful recovery of the ROI as compared to the other two schemes. Table 10 tabulates similar results for other sequences in our database.

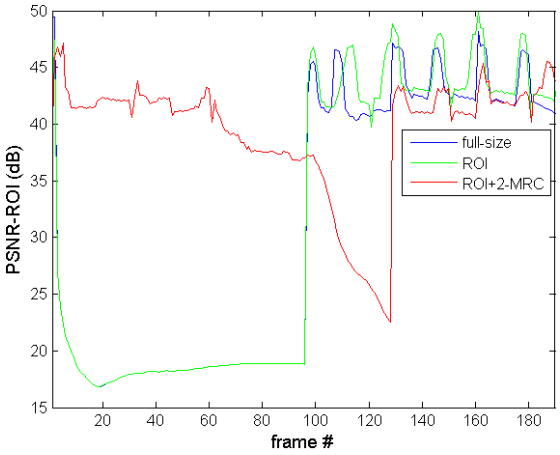
Figure 43 shows the effect of unequal protection of the ROI using the 2-MRC scheme. Frames 63 and 83 of the unimparied *Equation* sequence encoded uniformly, i.e. without any ROI can be seen in Figures 43a and 43b respectively. In case of burst losses multiple frames of the sequence are lost, and the decoder relies on the previous-frame-copy method, resulting in a freeze-frame effect as seen in Figures 43c



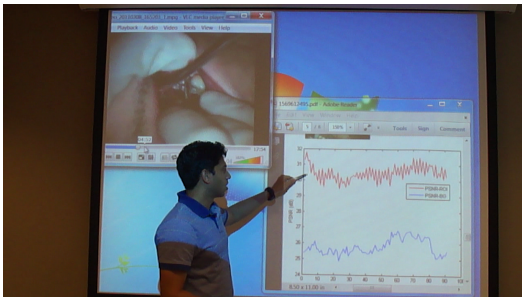
(a) BLL=0 (0 packets), 42.95 dB, 44.14 dB, 42.61 dB



(b) 1<BLL≤2 GOPs (1000 packets), 34.86 dB, 35.65 dB, 41.67 dB



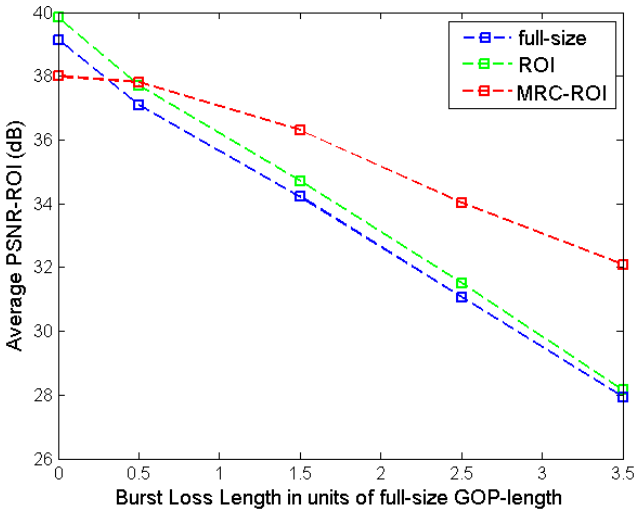
(c) 2<BLL≤3 GOPs (1700 packets), 31.10 dB, 31.62 dB, 39.33 dB



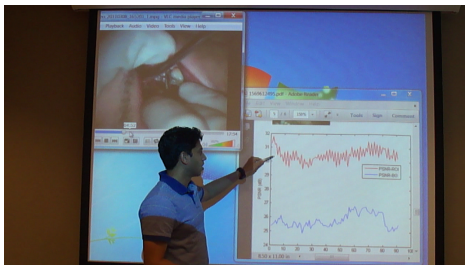
(d) The *Presentation* sequence

Figure 41: The *Presentation* sequence encoded with average bitrate of *2700 kbps* for the full-size, ROI, and the ROI+2-MRC schemes (Figure 40) schemes. The caption of each sub-figure indicates the burst loss length in units of GOP-length followed by the average PSNR-ROI of the full-size, ROI, and the ROI+2-MRC schemes.

and 43d. However, in the ROI+MRC scheme, the ROI is unequally protected using the MRC scheme. As a result, only the ROI of the sequence is delivered error-free as seen in Figure 43e. The BG in Figure 43e is as impaired as the entire frames of the full-size (Figure 43c) and the ROI schemes (Figure 43d).



(a) Average PSNR-ROI vs. BLL



(b) The *Presentation* sequence

Figure 42: Average PSNR-ROI vs. BLL in units of GOP-length for the *Presentation* sequence. Average PSNR-ROI was obtained by averaging the PSNR in the ROI over all frames of the sequence, and over all five loss traces, and over all four encoding bitrates.

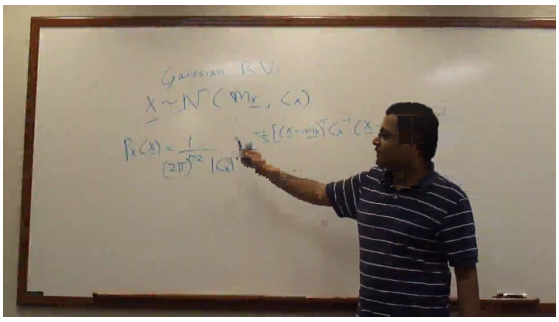
4.6 Summary

This chapter presented a flexible, interactive, and standard-compliant method to introduce Region of Interest support for enabling video streaming over low-bandwidth channels. An error-resilient video delivery scheme known as Multiple Representation Coding, which employed a novel GOP interleaver to temporally disperse multiple downsampled representations of the source video over a single transmitted stream, was also presented in this paper. Finally, it was demonstrated that unequal protection of the ROI-encoded sequences using the MRC scheme can facilitate a graceful recovery of the ROI from burst losses and signal loss intervals. The findings from this chapter have been submitted for publication in [56].

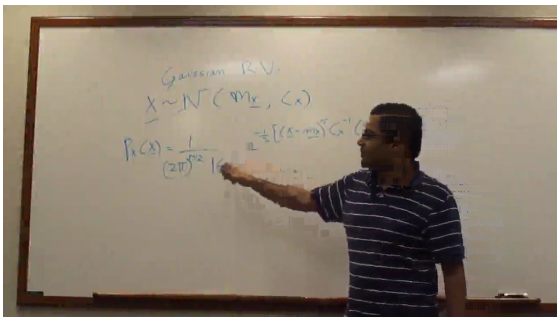
Table 10: Average PSNR-ROI for the full-size, ROI, and the ROI+2-MRC schemes for different test-sequences subject to various burst loss patterns.

Burst Loss Length	Speak (1700 kbps)			Equation (925 kbps)			Presentation (2700 kbps)		
	FS	**ROI**	**MRC +ROI**	**FS**	**ROI**	**MRC +ROI**	**FS**	**ROI**	**MRC +ROI**
BLL = 0	39.49	40.74	38.27	40.97	41.66	39.96	42.95	44.14	42.61
0<BLL≤1	29.62	30.31	38.21	37.60	38.13	39.59	40.00	40.87	42.19
1<BLL≤2	23.29	23.60	29.98	34.14	34.51	37.69	36.66	37.36	41.00
2<BLL≤3	22.55	22.83	24.42	29.70	29.92	33.99	32.97	33.47	37.44
3<BLL≤4	20.75	20.92	23.08	26.78	26.95	30.44	29.57	29.88	34.22

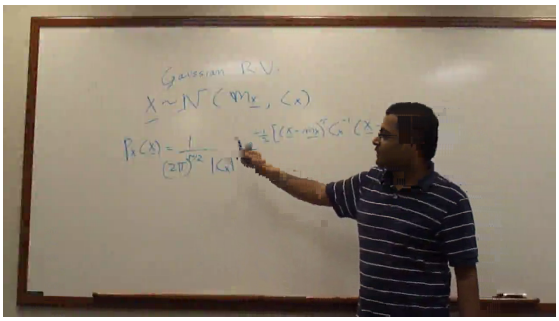
Burst Loss Length	ParkJoy (19 Mbps)			Sunflower (4.9 Mbps)		
	FS	**ROI**	**MRC +ROI**	**FS**	**ROI**	**MRC +ROI**
BLL = 0	30.09	32.65	28.80	42.39	43.70	41.80
0<BLL≤1	27.85	30.04	28.49	37.58	38.83	40.36
1<BLL≤2	25.20	26.92	26.77	33.35	34.23	39.17
2<BLL≤3	22.83	24.14	24.47	28.41	28.95	34.01
3<BLL≤4	20.63	21.55	22.00	23.18	23.51	30.64



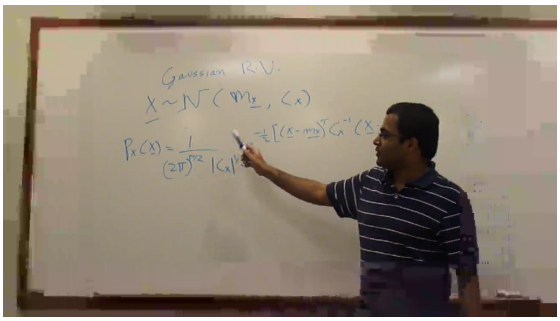
(a) Frame 63 of the unimpaired full-size sequence



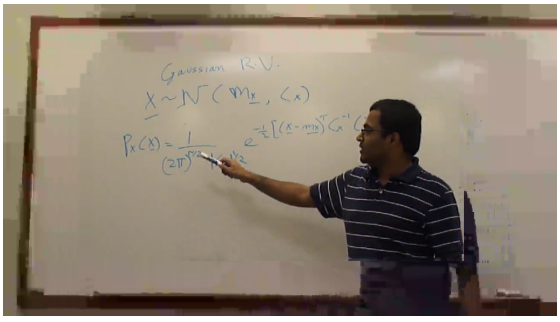
(b) Frame 83 of the unimpaired full-size sequence



(c) Frame 83 of the impaired full-size sequence



(d) Frame 83 of the impaired ROI-encoded sequence



(e) Frame 83 of the impaired ROI+2-MRC sequence

Figure 43: Frames of the *Equation* sequence encoded with average bitrate of *220 kbps* for the full-size, ROI, and the ROI+2-MRC schemes (Figure 40) schemes.

CHAPTER V

CONCLUSION AND FUTURE RESEARCH

Multimedia applications and services have matured to an extent where the end-user can now avail video communication in areas ranging from entertainment and shopping to telemedicine and tele-education. The end-user in the case of telemedicine applications is a medical expert responsible for remotely rendering a diagnosis based on the received medical imagery. In case of tele-education, the end-user is a remote learner participating in courses without physically attending lectures in the classroom. In most other cases, the end-user is simply a viewer initiating a mobile or wireless video streaming or video conferencing session. Regardless of the application, the end-user in each of these cases expects a near-perfect video quality for the entire duration of video communication. Unfortunately, providing a high quality video experience to the end-user is severely hampered by the lack of sufficient bandwidth and the presence of packet losses over the network. Furthermore, given the variety in the modality of the source video content, the diversity in the power and processing capabilities of the clients, and the heterogeneity in the nature of transmission channels, solutions to effectively address this challenge have to be both: content-dependent, and network-dependent. To that effect, this thesis presented an assortment of Application Specific Video Coding and Delivery algorithms to enable high-quality video communication over unfriendly environments. Relevant applications such as surgical telementoring, mobile and wireless video streaming, and distance learning were chosen to demonstrate the ability of the ASVCD toolkit in overcoming the unique challenges faced in rendering these different multimedia applications effective.

One of the applications investigated in this thesis was that of surgical telementoring. Surgical telementoring can effectively overcome distance, cost, and mobility issues, but may often be rendered ineffective due to lack of sufficient bandwidth. Scenarios necessitating delivery of live surgical procedures over low bandwidth wireless channels are very likely to occur in remote rural facilities and military field hospitals. In such settings, a meager bandwidth of about 128 to 200 kbps might be available for video communication. Current generation of video-coding standards such as H.264/AVC can encode videos at very low bitrate, but the quality of the compressed video is almost never diagnostically lossless. However, widespread deployment of surgical telementoring systems over bandwidth-scarce resources requires a definitive and compelling demonstration of the ability to deliver diagnostically lossless surgical videos to the remote mentor. To this end, this thesis proposed modifying an existing H.264 video encoder by introducing a flexible and interactive Region of Interest support. It was hypothesized that allocating more bits to the ROI as compared to the rest of the frame (also known as background) would facilitate the delivery of the surgical field with a visual quality considered appropriate for medical assessment. This hypothesis was supported by evaluations provided by multiple experts, who indicated enabling a high-quality ROI at the expense of a moderately degraded BG, is preferred over uniformly compressed surgical videos [53]. Further an ROI prototype which enabled the remote mentor to request live patient-video encoded with a flexible and interactive ROI was also implemented. Evaluation of the prototype by experts resulted in favorable reviews, again reaffirming the original hypothesis that ROI coding can prove very useful in enabling live transference of surgical videos over very low bandwidth channels.

Besides being stifled due to bandwidth insufficiencies, enabling a high quality video experience could also be hampered by the presence of packet losses over the underlying network. Presence of intermittent signal losses and bandwidth fluctuations

over commercial cellular data connections can significantly impair the user experience of mobile video consumers. In the worst case, users on a cellular network could experience complete loss of signal resulting in loss of multiple frames of a video sequence. However, existing approaches such as ARQ, FEC, FMO and others, face unavoidable limitations in the presence of burst losses or signal loss intervals. Although Multiple Description Coding can be an attractive option to combat a bursty channel, most MDC algorithms proposed in the literature assume a multipath communication capability, which cannot be guaranteed every time. This thesis proposed a scheme for error-resilient video streaming known as Multiple Representation Coding. In the MRC scheme, the frames of the source video are spatially or temporally downsampled to generate multiple independently decodable representations. These multiple representations are then temporally dispersed and transmitted as a single video stream using a novel GOP interleaver. The GOP-interleaver was designed to ensure that spatio-temporally co-located segments of the video sequence corresponding to different representations are not impaired by the same burst loss, even when transmitted over a single channel. Simulation results demonstrated that the MRC scheme can give a PSNR gain on the order of 2-4 dBs over the conventional full-size encoding and transmission of the video [52, 55]. Further, the inherent flexibility of the MRC scheme was employed by adaptively varying the GOP-length of the interleaved representations in response to the length of the burst loss over the network. This adaptive MRC approach resulted in improved error-recovery at the receiver, and gave a PSNR gain of around 2 dB over the non-adaptive MRC scheme in presence of sustained burst losses or signal loss intervals [54]. Furthermore, a comparative study of the proposed MRC scheme with Raptor-FEC demonstrated that MRC scheme can facilitate a more graceful recovery from impairments introduced by signal losses over the network. Finally, the effectiveness of the MRC scheme in facilitating a graceful recovery of HEVC sequences impaired by burst or signal losses was also demonstrated

in this thesis.

The third application presented in this thesis was that of enabling distance education or remote learning. Due to the recent initiatives by several education institutions, the ability to remotely access pre-taped or live lectures is becoming increasingly important. Unfortunately, due to the previously described heterogeneity in the network conditions and client capabilities, enabling a usable distance learning experience for the remote learners can be very challenging. It was reasoned in this thesis that delivering a uniformly-pristine video is not necessary for enabling distance learning applications. This thesis proposed delivering only the interesting regions of the frame with a higher quality, and with better error-resiliency to enable transmission over low-bitrate and error-prone channels. To that effect, the flexible and interactive ROI specified in the ASVCD toolkit, was employed to encode the critical regions within each frame at a higher quality as compared to the background. Further, the reconstruction fidelity of the high-quality ROI was improved by unequally protecting the ROI using the MRC scheme. Simulation results demonstrated that the proposed ROI+MRC strategy can enable a graceful recovery of the high-quality ROI in the presence of burst and signal losses [56].

To summarize, video as a communication modality is extremely demanding in terms of the bandwidth and reliability expected from the channel. This thesis proposed H.26x optimizations that addressed challenges associated with delivering video for applications such as telemedicine, entertainment and tele-education with an acceptable quality, over unreliable and bandwidth-constrained communication resources.

5.1 Future work

In spite of all the solutions proposed in this thesis, encoding and delivery of video content over unfriendly networks is certainly not a solved problem. Furthermore, the solutions presented in this thesis are not the only possible approaches to solve this

problem. Extensions to the work involving delivery of DL-quality videos for applications such as telementoring were already proposed in Chapter 2. It was shown in Section 2.5, that besides ROI encoding, preprocessing the source video to reduce the frame rate and/or the frame resolution could also help in encoding surgical videos with DL-quality without exceeding the target bitrate dictated by the channel. Other such methods that do not rely on the ROI coding technology could also be explored in the future. As mentioned in Chapter 2, application of methods such as ROI video coding or SVC for medical imagery requires a compelling evidence about their ability to achieve diagnostic losslessness. Measurement of DL quality cannot be performed using convenient metrics such as PSNR, but instead require extensive subjective evaluations by human experts. However, conducting such subjective experiments is often expensive and time-consuming. The presented research can be continued by investigating video quality metrics which can definitively measure the diagnostic quality without bearing the expense and complexity involved in conducting extensive subjective evaluations. The presented research can also be extended by incorporating expert feedback presented in Section 2.4 to develop a feature-laden ROI prototype with a multi-platform support for enabling ROI-based video coding. The ROI video-coding technology could possibly be commercialized by making it available in smartphone applications. Researchers involved in this project including the author, have already identified a market in the healthcare sector, and have submitted a VentureLab [101] proposal to facilitate a two-way, diagnostic quality, high-definition video communication over cellular networks.

Several extensions to the research involving MRC for error-resilient video delivery, are possible. For example, it was demonstrated in Chapter 3 that the GOP-length of interleaved representations can be adaptively varied to facilitate graceful recovery from burst or signal loss intervals. One direction of extending this work, would be to identify the optimal length of the interleaved segments, and also, the optimal number

of interleaved representations that can facilitate improved fidelity in channels with rapidly varying burst loss lengths. Further, one can assume models of the bursty channel (deterministic or statistical) and then recommend these MRC parameters more logically, thus reducing the burden of empirical experimentation. The performance of the MRC scheme can also be improved by investigating algorithms to improve the frame reconstruction at the receiver. Although a simple filter was employed in this thesis for interpolation of the missing pixels, improved reconstruction can be facilitated by using more sophisticated techniques such as new edge-directed interpolation [61] and soft-decision adaptive interpolation [116], at the cost of increased computational complexity. This complexity vs. quality trade-off could be worth exploring in the future. Finally, the authors believe that a multiple representation based architecture such as MRC or MDC, can find utility in enabling applications such as adaptive video streaming.

In closing, this thesis presented an assortment of content-dependent, and network-dependent solutions which found immediate utility in important and pervasive applications requiring high-quality video communication. Future avenues of research to improve and extend the ASVCD toolkit for enabling applications such as in-home patient care, and adaptive video streaming were also pointed out in this thesis. Future work could help fine tune the ASVCD toolkit, making its theory and application even more rigorous, and comparisons against prior art even more compelling. The author firmly believes that as the understanding of the research community about signal processing and communication theory keeps on improving, the solutions to address the numerous challenges presented in this thesis will also keep on refining, enabling high-quality video communications applications to become seamless, and pervasive.

List of Publications

- Sourabh Khire, Arturo Rodriguez, Scott Robertson, Nikil Jayant, " Error-resilient delivery of Region of Interest video using Multiple Representation Coding," International Conference on Acoustics, Speech, and Signal Processing, 2013 (accepted for publication).

- Sourabh Khire , Arturo Rodriguez, Nikil Jayant, "Temporal Dispersal of Multiple Representations for Error-Resilient Video Streaming," Asilomar Conference on Signals, Systems, and Computers, November 4-7, 2012, Pacific Grove, USA.

- Sourabh Khire , Scott Robertson, Nikil Jayant, Elena A Wood, Max E. Stachura, Col. Tamer Goksel, "Region-of-interest video coding for enabling surgical telementoring in low-bandwidth scenarios," Military Communications Conference, October 29 - November 1, 2012, Orlando, USA.

- Sourabh Khire , Arturo Rodriguez, Nikil Jayant, "Multiple representation coding for error-resilient video delivery," IEEE International Conference on Consumer Electronics - Berlin, Septemeber 3 - 5, 2012, Berlin, Germany.

- Sourabh Khire , Lee Cooper, Yuna Park, Alexis Carter, Nikil Jayant, Joel Saltz, "ZPEG: a hybrid DPCM-DCT based approach for compression of Z-stack images", 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August 28 - September 1, 2012, San Diego, USA.

- Sourabh Khire , Arturo Rodriguez, Nikil Jayant, "Adaptive GOP-length multiple representation coding for error-resilient video delivery", IET Image Processing Conference 2012, July 3 - 4, 2012, London, UK.

- Sourabh Khire , Scott Robertson, Nikil Jayant, Elena A Wood, Max E. Stachura,

Col. Tamer Goksel, "Diagnostically lossless delivery of surgical videos over low-bandwidth channels", submitted to T-ITB.

- Sourabh Khire , Arturo Rodriguez, Nikil Jayant, "Multiple representation Coding of streaming video for graceful recovery from signal losses", to be submitted to T-CE.

- Sourabh Khire , Lee Cooper, Saunya Williams, Nikil Jayant, Joel Saltz, "Initial Results with ZPEG: The Z-Stack Image Encoder", Pathology Informatics 2010, Sept 19 - 22, 2010, Boston, MA.

- Gee-Kung Chang, Joseph Long, Shu-Hao Fan, Cheng Liu, Arshad Chowdhury, Hung-Chang Chien, Sourabh Khire and Nikil Jayant, "Emerging Heterogeneous Optical Wireless Access Networks for Next Generation Telemedicine and Telehealth Applications", Opto-Electronics for Communications Conference, July 4-8 2011, Kaohsiung, Taiwan.

- Saunya M. Williams, Lee Cooper, Sourabh Khire , Joel H. Saltz, Nikil Jayant, "Compression Effects on Computer-Based Morphological Studies", Pathology Informatics 2010, Sept 19 - 22, 2010, Boston, MA.

- Arshad Chowdhury, Hung-Chang Chien, Sourabh Khire , Shu-Hao Fan, Nikil Jayant and Gee-Kung Chang, "Converged Broadband Optical and Wireless Communication Infrastructure for Next-Generation Telehealth", IEEE HealthCom'10, July 1-3 2010, Lyon, France.

- Arshad Chowdhury, Joshep Long, Hung-Chang Chien, Sourabh Khire , Nikil Jayant and Gee-Kung Chang, "Next-Generation E-Health Communication Infrastructure using Converged Super-Broadband Optical and Wireless Access System", IREHSS 2010 Conference, June 14, 2010, Montreal, Canada

- Sourabh Khire , Saunya Williams, Nikil Jayant, Alexis Carter, Uday Srinath, "One-stage and Two-stage Models for Compressing Pathology Image Slides", Advancing Practice Instruction and Innovation through Informatics Conference, Pittsburgh, Sept 20-23 2009.

- Sourabh Khire and Saibal Mukhopadhyay, "On improving the algorithmic robustness of a low-power FIR filter", IEEE International Conference on Computer Design, Oct 4-7 2009, Lake Tahoe, California.

- Chirag Pujara, Ashok Bhardwaj, Sourabh Khire and Vikram M. Gadre, "Secure watermarking in Fractional Wavelet domains", IETE Journal of Research, November-December 2007, Vol. 53, No. 6, pp. 573 - 580.

REFERENCES

[1] Agrafiotis, D., Bull, D., Canagarajah, N., and Kamnoonwatana, N., "Multiple Priority Region of Interest Coding with H.264," in *Proc. of IEEE International Conference on Image Processing*, pp. 53–56, October 2006.

[2] Antoniou, S., Franzen, J., Pointner, R., Antoniou, G., Bollmann, S., Koch, O., and Granderath, F., "A comprehensive review of telementoring applications in laparoscopic general surgery," *Surgical Endoscopy*, vol. 26, pp. 2111–2116, August 2012.

[3] Apostolopoulos, J. G., "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Proc. of Visual Communication and Image Processing*, pp. 392–409, January 2001.

[4] Arachchi, H. K., Fernando, W., Panchadcharam, S., and Weerakkody, W., "Unequal Error Protection Technique for ROI Based H.264 Video Coding," in *Canadian Conference on Electrical and Computer Engineering*, pp. 2033 –2036, may 2006.

[5] Baccichet, P., Zhu, X., and Girod, B., "Network-aware H.264/AVC region-of-interest coding for a multi-camera wireless surveillance network," in *Proc. of Picture Coding Symposium, (PCS-06)*, April 2006.

[6] Bagheri, M., Lotfi, T., Darabi, A., and Kasaei, S., "Content-based video coding for distance learning," in *IEEE International Symposium on Signal Processing and Information Technology*, December 2007.

[7] Bemardini, R., Durigon, M., Rinaldo, R., Celetto, L., and Vitali, A., "Polyphase spatial subsampling multiple description coding of video streams with H.264," in *Proc. of IEEE International Conf. on Image Processing*, p. 32133216, October 2004.

[8] Bjontegaard, G., *Calculation of average PSNR differences between RD-curves, document VCEG-M33*. ITU-T VCEG, April 2001.

[9] Bjontegaard, G., *Improvements of the BD-PSNR model, document VCEG-AI11*. ITU-T VCEG, July 2008.

[10] Boanca, C., Rafiq, A., Tamariz, F., Lavrentyev, V., Onisor, D., Flerov, E., Popescu, I., , and Merrell, R. C., "Remote video management for intraoperative consultation and surgical telepresence," *Telemedicine and e-Health*, vol. 13, pp. 603–608, October 2007.

[11] Bossen, F., *Common test conditions and software reference configurations*. ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), October 2010.

[12] Bouras, C., Kanakis, N., Kokkinos, V., and Papazois, A., "Embracing RaptorQ FEC in 3GPP multicast services," *Wireless Networks*, pp. 1–13, 2012.

[13] Bross, B., Han, W.-J., Ohm, J.-R., Sullivan, G. J., Wang, Y.-K., and Wiegand, T., *High Efficiency Video Coding (HEVC) text specification draft 10 (for FDIS & Consent), document JCTVC-L1003*. ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), January 2013.

[14] Cabral, J.E., J. and Kim, Y., "Multimedia systems for telemedicine and their communications requirements," *IEEE Communications Magazine*, vol. 34, pp. 20 –27, July 1996.

[15] Chai, D., Ngan, K., and Bouzerdoum, A., "Foreground/background bit allocation for region-of-interest coding," in *Proc. of IEEE International Conference on Image Processing*, vol. 2, pp. 923–926, September 2000.

[16] Chen, M. and Wei, G., "A novel hybrid ARQ algorithm for real-time video transport over wireless LAN," in *Proceedings on Personal, Indoor and Mobile Radio Communications, 2003*, vol. 3, pp. 2426–2430, September 2003.

[17] Chou, P., Mohr, A., Wang, A., and Mehrotra, S., "FEC and Pseudo-ARQ for receiver-driven layered multicast of audio and video," in *Proc. of IEEE Data Compression Conf.*, pp. 440–449, March 2000.

[18] Chu, Y. and Ganz, A., "A mobile teletrauma system using 3G networks," *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, pp. 456–462, December 2004.

[19] Cocosco, C. A., Netsch, T., Sngas, J., Bystrov, D., Niessen, W. J., and Viergever, M. A., "Automatic cardiac region-of-interest computation in cine 3D structural MRI," *International Congress Series*, vol. 1268, no. 0, pp. 1126 – 1131, 2004. Proceedings of the 18th International Congress and Exhibition Computer Assisted Radiology and Surgery.

[20] comScore, "More than 200 billion online videos viewed globally in october [online]." Available: `http://www.comscore.com/Insights/Press_Releases/2011/12/More_than_200_Billion_Online_Videos_Viewed_Globally_in_October`, December 2011. [accessed 1-January-2013].

[21] comScore, "comscore releases november 2012 u.s. online video rankings [online]." Available: `http://www.comscore.com/Insights/Press_Releases/2012/12/comScore_Releases_November_2012_U.S._Online_Video_Rankings`, December 2012. [accessed 1-January-2013].

[22] "Coursera [online]." Available: https://www.coursera.org/. [accessed 1-January-2013].

[23] Cubano, M., Poulose, B., Talamini, M., Stewart, R., Antosek, L., Lentz, R., Nibe, R., Kutka, M., and Mendoza-Sagaon, M., "Long distance telementoring. A novel tool for laparoscopy aboard the USS Abraham Lincoln," *Surgical endoscopy*, vol. 13, pp. 673–678, 1999.

[24] da Silva, V., McGregor, T., Rayman, R., and Luke, P. P., "Telementoring and telesurgery: Future or fiction? [online]." Robot Surgery, Seung Hyuk Baik (Ed.) ISBN: 978-953-7619-77-0, InTech, DOI: 10.5772/6895, Available: http://www.intechopen.com/articles/show/title/telementoring-and-telesurgery-future-or-fiction-, January 2010. [accessed 12-December-2012].

[25] Demartines, N., Mutter, D., Vix, M., Leroy, J., Glatz, D., Rsel, F., Harder, F., and Marescaux, J., "Assessment of telemedicine in surgical education and patient care," *Annals of Surgery*, vol. 231, no. 2, pp. 282–291, 2000.

[26] Dhondt, Y., Lambert, P., Notebaert, S., and Van de Walle, R., "Flexible macroblock ordering as a content adaptation tool in H.264/AVC," *Proc. of the SPIE*, vol. 6015, pp. 44–52, October 2005.

[27] Doukas, C. and Maglogiannis, I., "Adaptive transmission of medical image and video using scalable coding and context-aware wireless medical networks," *EURASIP Journal on Wireless Communications and Networking*, 2008. Article ID 428397, 12 pages, 2008, doi:10.1155/2008/428397.

[28] "edX [online]." Available: https://www.edx.org/. [accessed 1-January-2013].

[29] Ereso, A., Garcia, P., Tseng, E., Gauger, G., Kim, H., Dua, M., Victorino, G., and Guy, T., "Live transference of surgical subspecialty skills using telerobotic proctoring to remote general surgeons," *Journal of the American College of Surgeons*, vol. 211, pp. 400–411, September 2010.

[30] Fukaya, K., Noda, K., Tasaka, S., and Nunome, T., "The effect of audio-Video quality on learning effectiveness in distance learning over IP networks," in *IEEE Region 10 Conference TENCON 2010*, pp. 1444 –1449, November 2010.

[31] Gallant, M., Shirani, S., and Kossentini, F., "Standard-compliant multiple description video coding," in *Proc. of IEEE International Conf. on Image Processing*, p. 946949, October 2001.

[32] Gomez-Barquero, D. and Bria, A., "Forward Error Correction for File Delivery in DVB-H," in *IEEE 65th Vehicular Technology Conference*, pp. 2951 –2955, april 2007.

[33] Google, "Smartphone user study shows mobile movement under way [online]." "Available: http://googlemobileads.blogspot.com/2011/04/smartphone-user-study-shows-mobile.html", April 2011. [accessed 1-January-2012].

[34] Goyal, V. K., "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, pp. 74–93, September 2001.

[35] Group, I. N. W., "Raptor forward error correction scheme for object delivery [online]." Available: `http://tools.ietf.org/html/rfc5053`, October 2007. [accessed 1-January-2013].

[36] Guy, S., Herson, J., Knoth, B., Hayes, D., Grimsley, J., and Garcia, P., "Deployment of a surgical telementoring system in iraq: Lessons learned," in *15th Annual International Meeting and Exposition*, (San Antonio, TX), May 2010.

[37] *Video Codec for Audiovisual Services at p 384 64 kbit/s.* ITU-T Rec. H.261, 1990-1993.

[38] *Video Coding for Low Bit Rate Communication.* ITU-T Rec. H.263, Nov. 1995 (and subsequent editions).

[39] *Advanced Video Coding for Generic Audio-Visual Services.* ITU-T Rec. H.264 and ISO/IEC 14496-10 (AVC), ITU-T and ISO/IEC JTC 1,, May 2003 (and subsequent editions).

[40] Hannuksela, M., Wang, Y.-K., and Gabbouj, M., "Sub-picture: Roi coding and unequal error protection," in *Proc. International Conference on Image Processing*, vol. 3, pp. 537–540, June 2002.

[41] *Joint Call for Proposals on Video Compression Technology, document VCEG-AM91 of ITU-T Q6/16 and N1113 of JTC1/SC29/WG11.* ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC), January 2010.

[42] Hiatt, J., Shabot, M., Phillips, E., Haines, R., and Grant, T., "Acceptability of compressed video for remote surgical proctoring," *Archives of Surgery*, vol. 131, no. 4, pp. 396–400, 1996.

[43] Himawan, I., Song, W., and Tjondronegoro, D., "Impact of region-of-interest video coding on perceived quality in mobile video," in *Proc. IEEE International Conference on Multimedia and Expo*, pp. 79–84, July 2012.

[44] Hsiao, C.-W. and Tsai, W.-J., "Hybrid Multiple Description Coding Based on H.264," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 76–87, January 2010.

[45] Hu, D., "Image pre-processing for content-based image coding in a teleteaching system," in *Communication Technology Proceedings, 2000. WCC - ICCT 2000. International Conference on*, vol. 1, pp. 977 –980, 2000.

[46] INVODO, "Video statistics: The impact of videor [online]." "Available: http://www.invodo.com/html/resources/video-statistics/", 2010. [accessed 1-January-2012].

[47] JCT-VC, "Subversion Repository for the HEVC Test Model version HM7.0." Available: `https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-7.0/`. [Online].

[48] JOHANSON, M., "Adaptive forward error correction for real-time internet video," in *Proc. of the 13th Packet Video Workshop*, 2003.

[49] *Digital Compression and Coding of Continuous-Tone Still Images*. ITU-T and ISO/IEC JTC 1, Rec. T.81 and ISO/IEC 10 918-1 (JPEG), September 1992.

[50] "JSVM Software Manual [Online]." Available: http://evalsvc.googlecode.com/files/SoftwareManual.doc, June 2000. [accessed 1-January-2012].

[51] KARAMANOUKIAN, H. L., PANDE, R. U., PATEL, Y., FREEMAN, A. M., AOUKAR, P. S., and D'ANCONA, G., "Telerobotics, telesurgery, and telementoring," *Pediatric Endosurgery and Innovative Techniques*, vol. 7, pp. 421 – 425, December 2003.

[52] KHIRE, S., JAYANT, N., and RODRIGUEZ, A., "Multiple representation coding for error-resilient video delivery," in *Proc. IEEE International Conference on Consumer Electronics - Berlin*, pp. 6–10, September 2012.

[53] KHIRE, S., ROBERTSON, S., JAYANT, N., WOOD, E., STACHURA, M., and GOKSEL, C. T., "Region-of-interest video coding for enabling surgical telementoring in low-bandwidth scenarios," in *Proc. 2012 Military Communications Conference*, pp. 552–557, October 2012.

[54] KHIRE, S., RODRIGUEZ, A., and JAYANT, N., "Adaptive gop-length multiple representation coding for error-resilient video delivery," in *Proc. of IET Image Processing Conference 2012*, pp. 440–449, July 2012.

[55] KHIRE, S., RODRIGUEZ, A., and JAYANT, N., "Temporal dispersal of multiple representations for error-resilient video streaming," in *Asilomar Conference on Signals, Systems, and Computers*, November 2012.

[56] KHIRE, S., RODRIGUEZ, A. A., ROBERTSON, S., and JAYANT, N., "Error-resilient delivery of region-of-interest video using multiple representation coding," in *International Conference on Acoustics, Speech and Signal Processing 2013 (accepted for publication)*.

[57] KIM, D., YOO, S. K., PARK, I., CHOA, M., BAE, K., KIM, Y., and HEO, J., "A mobile telemedicine system for remote consultation in cases of acute stroke," *Journal of Telemedicine and Telecare*, vol. 15, no. 2, pp. 102–107, 2009.

[58] Kim, S. W., Kim, S. Y., Kim, S., and Heo, J., "Performance analysis of forward error correcting codes in IPTV," *IEEE Transactions on Consumer Electronics*, vol. 54, pp. 376–380, May 2008.

[59] Kumar, S., Xu, L., Mandal, M., and Panchanathan, S., "Error resiliency schemes in H.264/AVC standard," *Journal of Visual Communication and Image Representation*, vol. 17, pp. 425–450, April 2006.

[60] Li, M., Xu, Z., and Xu, Y., "A robust video coding scheme for mobile wireless communications," *Journal of Multimedia*, vol. 4, pp. 405–411, December 2009.

[61] Li, X. and Orchard, M., "New edge-directed interpolation," *Image Processing, IEEE Transactions on*, vol. 10, pp. 1521–1527, oct 2001.

[62] Liang, Y. J., Apostolopoulos, J. G., and Girod, B., "Analysis of packet loss for compressed video: does burst-length matter?," in *Proc. of IEEE International Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 684–687, April 2003.

[63] Lin, C.-W., Chang, Y.-J., and Chen, Y.-C., "Low-complexity face-assisted video coding," in *Proc. of IEEE International Conference on Image Processing*, vol. 2, pp. 207–210, sept. 2000.

[64] Luxtec, "Luxtec surgical video systems." Available: `http://www.transmedic.com.tw/doc/VideoCart_850383C.pdf`. [accessed 12-December-2012].

[65] Mao, S., Lin, S., Panwar, S. S., Wang, Y., and Celebi, E., "Video transport over ad hoc networks: multistream coding with multipath transport," *IEEE Journal on Selected Areas in Communications*, vol. 21, pp. 1721 – 1737, December 2003.

[66] Martini, M. and Mazzotti, M., "Quality driven wireless video transmission for medical applications," in *Proc. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3254–3257, September 2006.

[67] Merritt, L., "X264: A high performance H.264/AVC encoder." Available: `http://neuron2.net/library/avc/overview_x264_v8_5.pdf`. [Online].

[68] *Generic Coding of Moving Pictures and Associated Audio InformationPart 2: Video.* ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG 2 Video), ITU-T and ISO/IEC JTC 1, November 1994.

[69] *Coding of audio-visual objectsPart 2: Visual.* ISO/IEC 14496-2 (MPEG-4 Visual version 1), ISO/IEC JTC 1, 1999 (and subsequent editions).

[70] Nishimura, K., Kondo, T., and Aibara, R., "High quality video transfer system with dynamic redundancy of FEC over broadband network," in *IEEE Pacific Rim Conference on Communications, Computers and signal Processing*, vol. 2, pp. 903–906, August 2003.

[71] Nouri, N., Abraham, D., Moureaux, J. M., Dufaut, M., Hubert, J., and Perez, M., "Subjective MPEG2 compressed video quality assessment: Application to tele-surgery," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 764 –767, April 2010.

[72] Nystrom, M., Gibson, J., and Anderson, J., "Multiple description image coding using regions of interest," in *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, pp. 925–929, November 2007.

[73] Organization, V., "VLC media player [Online]." Available: `hwww.videolan.org/vlc/index.html`. [accessed 12-December-2012].

[74] P, T. and R, B., "Promises and pitfalls of the interactive television approach to teaching adult development and aging," *Educational Gerontology*, vol. 25, pp. 741–754, December 1999.

[75] Pappano, L., "The Year of the MOOC [Online]." Available: `http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html?pagewanted=all`, November 2012. [accessed 1-January-2013].

[76] Pedersen, P., Dickson, B., and Chakareski, J., "Telemedicine applications of mobile ultrasound," in *IEEE International Workshop on Multimedia Signal Processing*, pp. 1–6, oct. 2009.

[77] Podolsky, M., Mccanne, S., and Vetterli, M., "Soft ARQ for layered streaming media," *Journal of VLSI Signal Processing Systems*, vol. 27, pp. 81–97, February 2001.

[78] Poropatich, R., Lappan, C., and Lam, D., "Operational Use of U.S. Army Telemedicine Information Systems in Iraq and Afghanistan - Considerations for NATO Operations," in *Telemedicine for Trauma, Emergencies, and Disaster Management, Rifat Latifi (Ed.)*, pp. 173–182, Artech House, 2010.

[79] Ramshaw, B., Tucker, J., Duncan, T., Mason, E., and Lucas, G., "Laparoscopic herniorrhaphy: a review of 900 cases," *Surgical Endoscopy*, vol. 10, 1996.

[80] Rao, S. and Jayant, N., "Optimizing algorithms for region-of-interest video compression, with application to mobile telehealth," in *Multimedia and Expo, 2006 IEEE International Conference on*, pp. 513–516, July 2006.

[81] Rao, S., Jayant, N., Stachura, M., Astapova, E., and Pearson-Shaver, A., "Delivering diagnostic quality video over mobile wireless networks for telemedicine," *International Journal of Telemedicine and Applications*, 2009. Article ID 406753, 9 pages, 2009. doi:10.1155/2009/406753.

[82] Retailer, I., "Product videos raise purchase likelihood for stacks and stacks [online]." Available: `http://www.internetretailer.com/2011/03/07/product-videos-raise-purchase-likelihood-stacks-and-stacks`, March 2011. [accessed 1-January-2013].

[83] Rosser, J., Young, S., and Klonsky, J., "Telementoring: an application whose time has come," *Surgical endoscopy*, vol. 21, pp. 1458–1463, August 2007.

[84] Ryu, E.-S. and Jayant, N., "Home gateway for three-screen TV using H.264 SVC and Raptor FEC," *IEEE Transactions on Consumer Electronics*, vol. 57, pp. 1652–1660, November 2011.

[85] Schulam, P., Docimo, S., Saleh, W., Breitenbach, C., R.G., M., and Kavoussi, L., "Telesurgical mentoring. initial clinical experience," *Surgical Endoscopy*, vol. 11, pp. 1001 – 1005, October 1997.

[86] Schwarz, H., Marpe, D., and Wiegand, T., "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, pp. 1103 –1120, sept. 2007.

[87] Shokrollahi, A., "Raptor codes," *IEEE Transactions on Information Theory*, vol. 52, pp. 2551–2567, June 2006.

[88] Su, C., Yao, J. J., and Chen, H. H., "H.264/AVC-Based Multiple Description Coding Scheme," in *Proc. of IEEE International Conf. on Image Processing*, vol. 4, pp. 265–268, October 2007.

[89] Sullivan, G., Ohm, J., Han, W.-J., Wiegand, T., and Wiegand, T., "Overview of the high efficiency video coding (hevc) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, pp. 1649–1668, December 2012.

[90] Sun, M., Liu, Q., Xu, J., Kassam, A., Enos, S., Marchessault, R., Gilbert, G., and Sclabassi, R., "A smart video coding method for time lag reduction in telesurgery," in *Proc. of 25th Army Science Conference*, November 2006.

[91] Systems, C., "Cisco visual networking index: Forecast and methodology, 2011-2016 [online]." Available: `http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf`, 2012. [accessed 22-December-2012].

[92] Tan, T. K., Sullivan, G., and Wedi, T., "Recommended Simulation Conditions for Coding Efficiency Experiments." ITU-T SC16/Q6, 36th VCEG Meeting, San Diego, USA, 8 - 10 July 2008. Document VCEG-AJ10.

[93] Tan, W. and Zakhor, A., "Video multicast using layered FECs: And scalable compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 373–387, March 2001.

[94] Thomos, N., Argyropoulos, S., Boulgouris, N., and Strintzis, M., "Robust transmission of h.264/avc video using adaptive slice grouping and unequal error protection," in *Proc. IEEE International Conference on Multimedia and Expo*, pp. 593–596, July 2006.

[95] Tomlinson, M., Winders, R., Yarwood, P., and Roberts, S., "An evaluation of the application of video compression in distance learning and satellite communications," in *Digitally Compressed TV by Satellite, IEE Colloquium on*, pp. 7/1 –7/8, nov 1995.

[96] Tsai, M.-F., Chilamkurti, N., and Shieh, C.-K., "An adaptive packet and block length forward error correction for video streaming over wireless networks," *Wireless Personal Communications*, vol. 56, pp. 435–446, February 2011.

[97] Tsapatsoulis, N., Loizou, C., and Pattichis, C., "Region of Interest Video Coding for Low bit-rate Transmission of Carotid Ultrasound Videos over 3G Wireless Networks," in *Proc. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3717–3720, August 2007.

[98] Turkowski, K., "Graphics gems," ch. Filters for common resampling tasks, pp. 147–165, San Diego, CA, USA: Academic Press Professional, Inc., 1990.

[99] Valentine, D., "Distance learning: Promises, problems, and possibilities [online]." Available: `http://www.westga.edu/~distance/ojdla/fall53/valentine53.html`, 2002. [accessed 1-January-2013].

[100] Van Leuven, S., Van Schevensteen, K., Dams, T., and Schelkens, P., "An implementation of multiple region-of-interest models in H.264/AVC," in *Signal Processing for Image Enhancement and Multimedia Processing*, vol. 31 of *Multimedia Systems and Applications Series*, pp. 215–225, Springer US, 2008.

[101] "Venturelab at georgia tech [online]." Available: `http://venturelab.gatech.edu/`. [accessed 1-January-2013].

[102] Vetrivel, S., Suba, K., and Athisha, G., "An overview of H.26x series and its applications," *International Journal of Engineering Science and Technology*, vol. 2, pp. 4622–4631, September 2010.

[103] Wang, B., Kurose, J., Shenoy, P., and Towsley, D., "Multimedia streaming via tcp: An analytic performance study," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 4, pp. 16:1–16:22, May 2008.

[104] Wang, Y., Reibman, A., and Lin, S., "Multiple description coding for video delivery," in *Proceedings of the IEEE*, vol. 93, pp. 57–70, January 2005.

[105] Wang, Y. and Zhu, Q., "Error control and concealment for video communication: a review," *Proc. of the IEEE*, vol. 86, pp. 974 –997, May 1998.

[106] Webley, K., "Mooc brigade: Who is taking massive open online courses, and why? [online]." Available: `http://nation.time.com/2012/09/26/mooc-brigade-who-is-taking-massive-open-online-courses-and-why/`, September 2012. [accessed 1-January-2013].

[107] Wenger, S., Knorr, G. D., Ott, J., and Kossentini, F., "Error resilience support in H.263+," *IEEE Trans. Circuits Systems Video Technology*, vol. 8, pp. 867–877, November 1998.

[108] Wikipedia, "Distance education [online]." Available: `http://en.wikipedia.org/wiki/Distance_education`, 2013. [accessed 1-January-2013].

[109] Williams, S. M., Carter, A. B., and Jayant, N. S., "Diagnostically lossless compression of pathology image slides," in *Advancing Practice Instruction and Innovation through Informatics*, October 2008.

[110] Williams, S. M., Cooper, L., Khire, S., Saltz, J. H., and Jayant, N., "Compression effects on computer-based morphological studies," in *Pathology Informatics*, September 2010.

[111] Willis, B., "Distance education at a glance [online]." Available: `http://www.uiweb.uidaho.edu/eo/dist1.html`. [accessed 22-December-2012].

[112] Wong, A. and Kwok, Y., "On a region-of-interest based approach to robust wireless video transmission," in *Proc. of the 7th International Symposium on Parallel Architectures, Algorithms and Networks (I-SPAN 04)*, pp. 385–390, IEEE, May 2004.

[113] Yatawara, Y., Caldera, M., Kusuma, T., and Zepernick, H.-J., "Unequal error protection for ROI coded images over fading channels," in *Systems Communications, 2005. Proceedings*, pp. 111 – 115, August 2005.

[114] Yu, H., Lin, Z., and Pan, F., "Applications and improvement of H.264 in medical video compression," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, pp. 2707–2716, December 2005.

[115] Zhang, R., Regunathan, S. L., and Rose, K., "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Sel. Areas in Commun.*, vol. 18, pp. 966–976, June 2000.

[116] Zhang, X. and Wu, X., "Image Interpolation by Adaptive 2-D Autoregressive Modeling and Soft-Decision Estimation," *Image Processing, IEEE Transactions on*, vol. 17, pp. 887–896, june 2008.